



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

# Ανίχνευση Αγγειοδυσπλασιών σε Ροή Βίντεο Ενδοσκοπικής Κάψουλας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ  
ΤΟΥ  
**Ιωάννη Α. Βεζάκη**

Επιβλέπων: Δημήτριος Κουτσούρης  
Καθηγητής ΕΜΠ

Αθήνα, Δεκέμβριος 2018





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

# Ανίχνευση Αγγειοδυσπλασιών σε Ροή Βίντεο Ενδοσκοπικής Κάψουλας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ  
ΤΟΥ  
**Ιωάννη Α. Βεζάκη**

Επιβλέπων: Δημήτριος Κουτσούρης  
Καθηγητής ΕΜΠ

Εγκρίθηκε από την κάτωθι τριμελή επιτροπή την 20<sup>ή</sup> Δεκεμβρίου 2018.

---

Δημήτριος Κουτσούρης  
Καθηγητής

---

Γεώργιος Ματσόπουλος  
Καθηγητής

---

Ανδρέας Πολυδώρου  
Καθηγητής Ιατρικής ΕΚΠΑ

---

Ιωάννης Α. Βεζιάκης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών ΕΜΠ

Copyright © Ιωάννης Α. Βεζιάκης, 2018

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Στους γονείς μου.



# Ευχαριστίες

Με την περάτωση της παρούσας διπλωματικής εργασίας αισθάνομαι υποχρέωσή μου να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Δημήτριο Κουτσούρη για την ανάθεση του θέματος, την καθοδήγηση και την εμπιστοσύνη του. Επίσης θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου στην υπόλοιπη τριμελή επιτροπή, καθ. κ. Γεώργιο Ματσόπουλο και καθ. κ. Ανδρέα Πολυδώρου, καθώς και στη χειρουργική κλινική του Αρεταίειου νοσοκομείου που συνέβαλε επίσης, ανταποκρινόμενη σε όποια βοήθεια ζητήθηκε. Τέλος, αισθάνομαι την ανάγκη να εκφράσω ιδιαίτερες ευχαριστίες στον κ. Πέτρο Τουμπανιάρη για την πολύτιμη βοήθειά του χωρίς την οποία δε θα είχε ολοκληρωθεί αυτή η διπλωματική εργασία.





# Περίληψη

Η ανάπτυξη της ενδοσκοπικής κάψουλας αποτελεί μία ριζική καινοτομία η οποία άλλαξε εντελώς τον τρόπο με τον οποίο γίνεται η εξέταση του λεπτού εντέρου. Καταπίνοντας την ιδιαίτερα μικρή αυτή κάψουλα, σε μορφή και μέγεθος χαπιού, εικόνες από το γαστρεντερικό σωλήνα του ασθενούς αποστέλλονται σε πραγματικό χρόνο σε μια εξωτερική συσκευή καταγραφής. Ωστόσο, λόγω του μήκους του εντέρου, η παρακολούθηση ενός τέτοιου βίντεο αποτελεί μια επίπονη και χρονοβόρα διαδικασία.

Από την στιγμή που η κάψουλα έκανε την εμφάνισή της στην αγορά, πολλές μελέτες έχουν δημοσιευθεί οι οποίες ασχολούνται με την βελτίωση της εικόνας, την επιτάχυνση της εξέτασης, και την αυτόματη ανίχνευση παθολογιών. Ειδικά για την ανίχνευση παθολογιών, λίγα εργαλεία υπάρχουν αυτή τη στιγμή τα οποία να μπορούν να επιτύχουν κάτι τέτοιο, ενώ η χρησιμότητά τους είναι, μάλλον, αμφίβολη.

Σε αυτή τη διπλωματική εργασία, παρουσιάζεται μία νέα μέθοδος ανίχνευσης μιας πολύ κοινής, ιδιαίτερα σε ηλικιωμένους ανθρώπους, παθολογίας, των αγγειοδυσπλασιών. Για την πλήρη κατανόηση της μεθόδου αλλά και της χρησιμότητάς της, γίνεται μια εισαγωγή στην ενδοσκόπηση, ενώ στη συνέχεια, παρουσίαση της ενδοσκοπικής κάψουλας και των αγγειοδυσπλασιών. Στο κομμάτι της όρασης υπολογιστών, αναλύεται τι είναι μια ψηφιακή εικόνα, πώς αυτή μπορεί να αναλυθεί από έναν υπολογιστή, ενώ επεξηγούνται επίσης οι χρωματικοί χώροι. Στη συνέχεια, γίνεται ανάλυση των κλασικών και συνελκτικών νευρωνικών δικτύων. Τέλος, παρουσιάζεται βήμα-βήμα η λειτουργία του αλγορίθμου.

Λέξεις-κλειδιά: Ενδοσκοπική Κάψουλα, Αγγειοδυσπλασίες, Όραση Υπολογιστών, Νευρωνικά Δίκτυα, Διάγνωση Υποβοηθούμενη από Υπολογιστή (CAD)



# Abstract

The development of capsule endoscopy is a radical innovation that revolutionized the way with which the examination of the small bowel is performed. By swallowing this small, pill-like in both shape and size, capsule, images of the patient's gastrointestinal tract are transmitted in real time to an external recorder. However, due to the small bowel's length, the inspection of the resulting video is a laborious and time-consuming process.

Since the capsule first emerged in the market, many studies have been published on improving the image, accelerating the examination and automatically diagnosing pathologies. At the moment, not many tools exist that can perform an automatic pathology diagnosis, and the usefulness of those that do exist is rather ambiguous.

In this diploma thesis, a novel method for automatic detection of angioectasias is presented - a very common, especially among older people, pathology. In order to make this method and its usefulness clearly understood, first an introduction to endoscopy is made. Then, capsule endoscopy and angioectasias are presented. On the topic of computer vision, it is explained what a digital image is, and how it can be analyzed by a computer. Color spaces are also explained. Following that, conventional and convolutional neural networks are described. Finally, the algorithm is presented step-by-step.

Keywords: Capsule Endoscopy, Angioectasias, Computer Vision, Neural Networks, Computer Aided Diagnosis (CAD)

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>17</b>
1.1	Ενδοσκόπηση . . . . .	17
1.2	Η ενδοσκοπική κάψουλα . . . . .	18
1.2.1	Προετοιμασία ασθενή . . . . .	19
1.2.2	Διαδικασία . . . . .	20
1.2.3	Ανάλυση . . . . .	20
1.2.4	Ενδείξεις . . . . .	21
1.2.5	Αντενδείξεις . . . . .	22
1.2.6	Δομικά μέρη . . . . .	22
<b>2</b>	<b>Αγγειοδυσπλασίες</b>	<b>25</b>
2.1	Παθοφυσιολογία . . . . .	25
2.2	Παθογένεση . . . . .	26
2.3	Διάγνωση . . . . .	27
2.4	Θεραπεία . . . . .	28
<b>3</b>	<b>Ανάλυση εικόνας</b>	<b>31</b>
3.1	Αρχεία εικόνας . . . . .	32
3.1.1	Δειγματοληψία . . . . .	33
3.1.2	Κβάντιση . . . . .	33
3.2	Γειτονικές πράξεις (Neighborhood operations) . . . . .	34
3.2.1	Γενικός ορισμός . . . . .	35
3.2.2	Γραμμικό φιλτράρισμα . . . . .	36
3.2.3	Συνέλιξη . . . . .	36
3.2.4	Παραδείγματα . . . . .	38
3.3	Χρωματικοί χώροι . . . . .	40
3.3.1	Ανθρώπινη αντίληψη . . . . .	42
3.3.2	Περιγραφή του χρώματος . . . . .	42
3.3.3	Γραμμικοί χώροι χρώματος . . . . .	42
3.3.4	Μη-γραμμικοί χώροι χρώματος . . . . .	45
<b>4</b>	<b>Νευρωνικά Δίκτυα</b>	<b>49</b>
4.1	Δομή . . . . .	49
4.1.1	Το μοντέλο ενός νευρώνα . . . . .	52

4.2	Εκπαίδευση νευρωνικών δικτύων . . . . .	54
4.2.1	Κανόνες εκμάθησης . . . . .	54
4.2.2	Αλγόριθμος απότομης καθόδου . . . . .	55
4.2.3	Backpropagation . . . . .	58
<b>5</b>	<b>Ανίχνευση αγγειοδυσπλασιών</b>	<b>61</b>
5.1	State-of-the-art μέθοδοι . . . . .	61
5.1.1	Ανίχνευση αιμορραγιών . . . . .	62
5.1.2	Ανίχνευση αγγειοδυσπλασιών . . . . .	62
5.1.3	Χρόνος εξέτασης . . . . .	62
5.2	Άλλες έρευνες και μέθοδοι εντοπισμού παθολογιών . . . . .	62
5.3	Περιοχές ενδιαφέροντος . . . . .	64
5.3.1	Χώρος CIE Lab . . . . .	64
5.3.2	K-means clustering . . . . .	65
5.3.3	Contrast, Gaussian filtering . . . . .	68
5.3.4	Speeded-Up Robust Feature Detection . . . . .	69
5.4	CNN κατηγοριοποίηση . . . . .	72
5.4.1	Συνελικτικά νευρωνικά δίκτυα . . . . .	74
5.4.2	Εκπαίδευση . . . . .	76
5.4.3	Εφαρμογή στο πρόβλημα αναγνώρισης αγγειοδυσπλασιών . . . . .	78
<b>6</b>	<b>Αποτελέσματα</b>	<b>81</b>
<b>7</b>	<b>Συμπεράσματα</b>	<b>85</b>

# Κατάλογος Σχημάτων

1.1	Μια ενδοσκοπική κάψουλα σε πραγματικές διαστάσεις . . . . .	19
1.2	Block διάγραμμα ενδοσκοπικής κάψουλας . . . . .	23
2.1	Χαρακτηριστική εικόνα αγγειοδυσπλασίας από ενδοσκοπική κάψουλα	26
3.1	Τμήματα κάμερας με φακό και σχηματισμός εικόνας . . . . .	32
3.2	Παράδειγμα κβάντισης μονόχρωμης εικόνας . . . . .	34
3.3	Διακριτή δισδιάστατη συνέλιξη. . . . .	37
3.4	Η γκαουσιανή κατανομή με $\sigma = 0.5$ και $\sigma = 0.75$ . . . . .	39
3.5	Παράδειγμα εφαρμογής φίλτρων σε μια εικόνα σκύλου. . . . .	41
3.6	Το ορατό φάσμα φωτός. . . . .	42
3.7	Οι RGB χρωματικές συνιστώσες. . . . .	43
3.8	Συναρτήσεις χρωματικού ταιριάσματος για τα RGB . . . . .	45
3.9	Το χρωματικό διάγραμμα και το τρίγωνο RGB . . . . .	46
3.10	Ο χρωματικός χώρος HSV . . . . .	47
3.11	Ελλείψεις MacAdam . . . . .	48
4.1	Σχηματική αναπαράσταση ενός νευρώνα . . . . .	50
4.2	Παράδειγμα 4 εισόδων σε ένα νευρωνικό δίκτυο . . . . .	50
4.3	Κρυφά επίπεδα σε ένα νευρωνικό δίκτυο . . . . .	51
4.4	Μοντέλο πλήρους νευρωνικού δικτύου . . . . .	51
4.5	Η σιγμοειδής συνάρτηση με $\alpha = 10$ . . . . .	53
4.6	Το μοντέλο ενός νευρώνα . . . . .	54
4.7	Γραφική παράσταση τρισδιάστατης συνάρτησης . . . . .	56
4.8	Γραφική απεικόνιση της απότομης καθόδου . . . . .	57
5.1	Εικόνα από CE βίντεο με οριακά εμφανή αγγειοδυσπλασία . . . . .	64
5.2	Τα 3 κανάλια της Lab εικόνας. . . . .	65
5.3	Εφαρμογή της μάσκας στην εικόνα . . . . .	68
5.4	Μεταβολή αντίθεσης της εικόνας και εφαρμογή γκαουσιανού φίλτρου	69
5.5	Προσέγγιση με box filters . . . . .	70
5.6	Υπολογισμός του ολοκληρωτικού αθροίσματος $S'$ . . . . .	71
5.7	Ανίχνευση σημείων ενδιαφέροντος . . . . .	72
5.8	Παράδειγμα ανίχνευσης περιοχής ενδιαφέροντος . . . . .	73
5.9	Επιπλέον παράδειγμα ανίχνευσης περιοχής ενδιαφέροντος . . . . .	73

5.10	Η επίδραση του convolutional και του pooling επιπέδου στην είσοδο	76
5.11	Η επίδραση της κανονικοποίησης σε ένα σύνολο δεδομένων. . . . .	79
5.12	Δείγμα εικόνων από το σύνολο δεδομένων . . . . .	80
6.1	Δείγμα από τις αγγειοδυσπλασίες οι οποίες βρέθηκαν με χρήση του αλγορίθμου στα τρία CE βίντεο. . . . .	83





# Κεφάλαιο 1

## Εισαγωγή

Η παρατήρηση καθώς και η απεικόνιση του εσωτερικού του ανθρώπινου σώματος αποτελεί ένα βασικό κομμάτι της ιατρικής τόσο για την κατανόηση όσο και για τη διάγνωση παθολογιών. Περίπου 1000 χρόνια πριν, ο Άραβας ιατρός Albukasim (916 - 1013 μ.Χ.) φέρεται να είναι ο πρώτος ο οποίος χρησιμοποίησε ανακλώμενο φως για την παρατήρηση ενός εσωτερικού οργάνου - της μήτρας [1].

Σήμερα διαθέτουμε μία πληθώρα εργαλείων για την ενδοσκόπηση του σώματός μας. Ιδιαίτερα για το γαστρεντερικό σωλήνα, έχουμε πρόσβαση τόσο στον οισοφάγο όσο και στο στομάχι με τη χρήση εύκαμπτων γαστροσκοπίων. Αντίστοιχα, με τη χρήση κολonosκοπίων είναι δυνατή η πρόσβαση στο παχύ έντερο. Μέχρι πρόσφατα όμως, λόγω της θέσης, του μήκους, και της πολυπλοκότητάς του, το λεπτό έντερο παρέμενε απροσπέλαστο.

Μία λύση σε αυτό το πρόβλημα δόθηκε με τη χρήση της ενδοσκοπικής κάψουλας (capsule endoscopy). Η ενδοσκοπική κάψουλα είναι μία μικροσκοπική κάμερα, σε μορφή και μέγεθος χαπιού, η οποία μπορεί έπειτα από κατάποση να αποστείλει σε πραγματικό χρόνο εικόνα βίντεο. Συνεπώς η διάγνωση παθολογιών στο λεπτό έντερο συνίσταται στην παρακολούθηση, από τον ιατρό, ολόκληρης της διαδρομής της κάψουλας μέσα σε αυτό - μία χρονοβόρα και ανιαρή διαδικασία, με αρκετά μεγάλη πιθανότητα λάθους ή παράλειψης [2]. Σήμερα, ελάχιστα εργαλεία υπάρχουν για την βελτίωση της διάγνωσης του βίντεο, τόσο στο ποσοστό επιτυχίας όσο και στην ταχύτητά της.

### 1.1 Ενδοσκόπηση

Τα τελευταία χρόνια έχει αναθεωρηθεί σε σημαντικό βαθμό η αντίληψη για την αντιμετώπιση των πιο συνηθισμένων χειρουργικών παθήσεων με τις κλασικές μεθόδους. Με τη χρήση εύκαμπτων ενδοσκοπίων καθώς και λαπαροσκοπίων τα οποία διέρχονται από μικρές τομές ή διά φυσικών οπών του σώματος είναι δυνατή η διάγνωση και οι θεραπευτικές επεμβάσεις με το μικρότερο δυνατό χειρουργικό τραύμα. Οι τεχνικές αυτές περιγράφονται με το δημοφιλή όρο «ελάχιστα επεμβατική χειρουργική

(MIS<sup>1</sup>)».

Η επεμβατική ενδοσκόπηση του πεπτικού αποτελεί σήμερα έναν αυτοτελή κλάδο της Ιατρικής, ο οποίος αντιπροσωπεύει ένα από τα πλέον εντυπωσιακά επιτεύγματα της τεχνολογικής εξέλιξης των τελευταίων δεκαετιών. Αυτή η ραγδαία ανάπτυξη της ενισχύθηκε επιπλέον από την εισαγωγή βιντεοενδοσκοπιών υψηλής ευκρίνειας και της βιοϊατρικής τεχνολογίας γενικότερα.

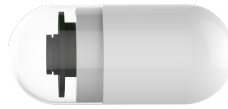
## 1.2 Η ενδοσκοπική κάψουλα

Η απεικόνιση στη συμβατική ενδοσκόπηση επιτυγχάνεται με τη χρήση οπτικών ινών, οι οποίες περικλειόμενες από έναν εύκαμπτο σωλήνα εισέρχονται σε εσωτερικές κοιλότητες όπως είναι το στομάχι και το παχύ έντερο, μεταφέροντας αμφίδρομα φως. Πρόσφατα, οι οπτικές ίνες έχουν αντικατασταθεί από καλώδια τα οποία τροφοδοτούν τη λειτουργία CCD καμερών καθώς και LED πηγών φωτός τοποθετημένα στο άκρο του εύκαμπτου αυτού σωλήνα - του ενδοσκοπίου. Η ενδοσκόπηση με αυτόν τον τρόπο είναι ιδιαίτερα άβολη για τον ασθενή, τόσο λόγω του μήκους των ενδοσκοπίων όσο και του πάχους τους. Ιδιαίτερα για την εξέταση του λεπτού εντέρου, τα συμβατικά ενδοσκόπια αποδεικνύονται άκρως προβληματικά, όχι μόνο λόγω του εξαιρετικά μεγάλου μήκους του οργάνου σε συνδυασμό με το περιορισμένο μήκος του ενδοσκοπίου, αλλά και της δυσφορίας του ασθενή κατά τη διάρκεια της εξέτασης. Ως συνέπεια, η εξέταση του λεπτού εντέρου με ένα συμβατικό ενδοσκόπιο περιορίζεται αναγκαστικά μόνο μέχρι το δωδεκαδάκτυλο. Πρόσφατα, καινούργιες τεχνικές έχουν αναπτυχθεί όπως αυτή του εντεροσκοπίου (double-balloon endoscopy - DBE) οι οποίες αν και μπορούν να εξετάσουν το σύνολο του λεπτού εντέρου παραμένουν ιδιαίτερα άβολες για τον ασθενή και απαιτητικές για τον ιατρό [3].

Η ανάπτυξη, συνεπώς, της ενδοσκοπικής κάψουλας είναι ιδιαίτερα σημαντική καθώς καλύπτει τις κλινικές ανάγκες για βελτίωση των υπαρχόντων μεθόδων εξέτασης. Η πρώτη ενδοσκοπική κάψουλα η οποία εγκρίθηκε από την FDA<sup>2</sup> των ΗΠΑ το 2001, ήταν της εταιρίας Given<sup>®</sup> Imaging με το, μάλλον ατυχές, όνομα Mouth-to-Anus (M2A<sup>®</sup>). Έκτοτε τουλάχιστον 4 ακόμη κάψουλες παγκοσμίως έχουν κάνει την είσοδό τους στην αγορά. Αν και διαφέρουν σημαντικά σε αρκετές τεχνολογικές προδιαγραφές, κοινό χαρακτηριστικό όλων είναι η ευκολία με την οποία γίνεται η εξέταση για τον ασθενή, αφού πρόκειται για κάψουλες τόσο μικρές σε μέγεθος (περίπου 11x30 mm) ώστε η κατάποσή τους να είναι ιδιαίτερα εύκολη. Από τις κάψουλες οι οποίες υπάρχουν σήμερα στην αγορά, αυτή η οποία χρησιμοποιείται περισσότερο στις δημοσιευμένες έρευνες είναι η PillCam<sup>®</sup> της Given<sup>®</sup> Imaging. Τα κύρια μέρη από τα οποία αποτελείται είναι ο αισθητήρας εικόνας (συνήθως CMOS τεχνολογίας, αν και υπάρχουν κάψουλες οι οποίες χρησιμοποιούν CCD αισθητήρες), η LED πηγή φωτός, το ολοκληρωμένο κύκλωμα ειδικών εφαρμογών (ASIC - application specific integrated circuit) και η μπαταρία.

<sup>1</sup> Minimally Invasive Surgery

<sup>2</sup> Food and Drug Administration



Σχήμα 1.1: Μια ενδοσκοπική κάψουλα σε πραγματικές διαστάσεις

Κατόπιν της κατάποσης από τον ασθενή, η κάμερα προωθείται αυτόματα μέσα από το σύνολο του γαστρεντερικού σωλήνα με το φυσικό εντερικό περισταλτισμό, και καταγράφει την διαδρομή αποστέλλοντας τις εικόνες για αποθήκευση σε μία εξωτερική φορητή συσκευή την οποία πρέπει να φέρει μαζί του ο ασθενής καθ' όλη τη διάρκεια της εξέτασης. Από αυτή τη συσκευή λαμβάνεται αργότερα το τελικό αρχείο βίντεο το οποίο μπορεί και να παρακολουθήσει ο ιατρός σε έναν υπολογιστή για την πραγματοποίηση της διάγνωσης. Με ταχύτητα λήψης 2 εικόνων το δευτερόλεπτο, το βίντεο αυτό περιλαμβάνει περίπου 86.400 εικόνες όταν πρόκειται για μία ολοκληρωμένη εξέταση. Μετά το πέρας της εξέτασης η κάψουλα αποβάλλεται από το έντερο διά της φυσικής οδού.

### **1.2.1 Προετοιμασία ασθενή**

Προκειμένου να εξασφαλισθεί μια σαφής και καθαρή εικόνα στο βίντεο της ενδοσκοπικής κάψουλας, ο ασθενής πρέπει να παραμείνει νηστικός για 12 ώρες πριν την εξέταση [4]. Ωστόσο, λόγω φυσαλίδων αέρα, του εντερικού υγρού και των εκκρίσεων των χοληφόρων και του παγκρέατος που προέρχονται από την κύρια δωδεκαδακτυλική θηλή, η εικόνα μπορεί να μην είναι καθαρή [5]. Επιπλέον, η περιορισμένη διάρκεια ζωής της μπαταρίας μπορεί να παρεμποδίσει την ολοκλήρωση της εξέτασης σε ασθενείς με καθυστερημένη γαστρική κένωση, πράγμα που απαιτεί τη χρήση επιπρόσθετης προετοιμασίας του λεπτού εντέρου [6].

Ωστόσο δεν είναι όλοι οι ασθενείς κατάλληλοι για επιπρόσθετο καθαρισμό του λεπτού εντέρου. Οι οδηγίες για την ασφαλή συνταγογράφηση και την εκ του στόματος χορήγηση προληπτικών παραγόντων καθαρισμού του εντέρου αναφέρει ότι η προετοιμασία του λεπτού εντέρου αντενδείκνυται σε ασθενείς με γαστρεντερική απόφραξη, διάτρηση, έλκος, ειλεό, γαστρική κατακράτηση ή φλεγμονώδεις παθήσεις των εντέρων, σε ασθενείς με μειωμένο επίπεδο συνείδησης, διαταραχές κατάποσης, υπερευαίσθησία στον χρησιμοποιούμενο παράγοντα και σε ασθενείς με ειλεοστομία [7]. Ιδιαίτερη προσοχή για τη χρήση παραγόντων καθαρισμού του λεπτού εντέρου χρειάζεται επίσης σε ασθενείς με χρόνια νεφρική νόσο ή που υποβάλλονται σε αιμοκάθαρση, σε ασθενείς με μεταμόσχευση νεφρού, συμφορητική καρδιακή ανεπάρκεια, κίρρωση του ήπατος ή ασκίτη, και σε ασθενείς που παίρνουν αναστολείς της ρενίνης-αγγειοτενσίνης, διουρητικά ή μη στεροειδή αντιφλεγμονώδη φάρμακα. Σε αυτούς τους ανθρώπους, η χρησιμότητα του καθαρισμού του λεπτού εντέρου πρέπει να επανεξετασθεί και η επιλογή του παράγοντα καθαρισμού έχει μεγάλη σημασία [5]. Επιπλέον, συνιστάται αποχή των ασθενών από το κάπνισμα, καθώς η κατάποση καπνού προσδίδει σκούρα χροιά στο γαστρικό βλεννογόνο και τροποποιεί την

κινητικότητα του εντέρου [8].

### **1.2.2 Διαδικασία**

Κατόπιν της προετοιμασίας του εντέρου, οκτώ συστοιχίες αισθητήρων επικολλούνται στη κοιλιακή χώρα του ασθενούς, και μια ειδικά σχεδιασμένη ζώνη δένεται γύρω από τη μέση του η οποία φέρει κατάλληλες θήκες για τη συσκευή καταγραφής [9]. Η συσκευή καταγραφής δεδομένων προσαρτάται στη ζώνη. Η κατάποση της κάψουλας γίνεται με ένα ποτήρι νερό, το οποίο μπορεί να περιέχει σταγόνες σιμεθικόνης για τη μείωση των φυσαλίδων εντός του εντερικού αυλού [8], ενώ στη συνέχεια απαιτείται περιορισμός των υγρών για έως και 2 ώρες μετά την κατάποση. Μετά την πάροδο 4 ωρών, η νηστεία μπορεί να σταματήσει. Οι καθημερινές δραστηριότητες δεν χρειάζεται να διακόπτονται κατά τη διάρκεια της εξέτασης, με τον περιορισμό ότι ο ασθενής θα πρέπει να μείνει μακριά από οποιαδήποτε πηγή ηλεκτρομαγνητικού πεδίου (π.χ. μαγνητικό τομογράφο, πομπούς ραδιοφωνικών σταθμών), για να μην παρατηρηθεί απώλεια και αλλοίωση της ποιότητας των ενδοσκοπικών εικόνων. Ο ασθενής συμβουλεύεται να είναι περιπατητικός, ιδιαίτερα τις πρώτες 1 έως 2 ώρες μετά την κατάποση της κάψουλας.

Η κατάποση της κάψουλας πρέπει απαραίτητως να ακολουθηθεί από παρακολούθηση, σε πραγματικό χρόνο, της εικόνας για την πρώτη ώρα, ώστε να επιβεβαιωθεί το πέρασμά της από το στομάχι. Αν δεν γίνει αυτό, πραγματοποιείται γαστροσκόπηση από τον ιατρό για την εναπόθεσή της στο δωδεκαδάκτυλο [5].

Οι συστοιχίες αισθητήρων και η ζώνη αφαιρούνται μόλις η κάψουλα προωθηθεί στο παχύ έντερο, αφού επιβεβαιωθεί η τοποθεσία της με την προβολή της εικόνας σε πραγματικό χρόνο, ή μετά το πέρας της διάρκειας ζωής της μπαταρίας. Οι εικόνες μπορούν να μεταφορτωθούν από τη συσκευή εγγραφής στο σταθμό εργασίας. Ο σταθμός εργασίας είναι ένας ηλεκτρονικός υπολογιστής με ειδικό λογισμικό, στον οποίο μεταφέρονται τα ψηφιακά δεδομένα που έχουν αποθηκευτεί στη συσκευή καταγραφής, όπου αναλύονται και αναπαράγονται οι εικόνες υπό μορφή έγχρωμου βίντεο. Η κάψουλα στη συνέχεια προωθείται φυσικά με τον περισταλτισμό του εντέρου και συνήθως αποβάλλεται εντός 24 έως 72 ωρών διά της φυσικής οδού [5]. Αν μετά την παρέλευση τριών ημερών ο ασθενής δεν επιβεβαιώσει την αποβολή της κάψουλας, θα πρέπει να γίνει ακτινογραφία κοιλίας, καθώς υπάρχει το ενδεχόμενο μόνιμης κατακράτησης μέσα στον εντερικό αυλό, ακόμη και επί απουσίας κοιλιακών συμπτωμάτων.

### **1.2.3 Ανάλυση**

Έπειτα από τη μεταφόρτωση των δεδομένων από τη συσκευή εγγραφής στο σταθμό εργασίας, οι εικόνες μπορούν να εξετασθούν από γαστρεντερολόγους χρησιμοποιώντας το κατάλληλο λογισμικό. Ο χρόνος εξέτασης μπορεί να ποικίλει, αλλά συνήθως είναι περίπου 60-120 λεπτά [10]. Νέες μελέτες, ωστόσο, αναφέρουν πως με τη χρήση των σύγχρονων τεχνικών απεικόνισης και επεξεργασίας εικόνας, ο χρόνος

αυτός μπορεί να μειωθεί σημαντικά [11]. Λόγω, όμως, της χρονοβόρας αυτής διαδικασίας, μία λύση σε αυτό το πρόβλημα είναι η εκπαίδευση μη-ιατρικού προσωπικού για την προανάγνωση των εικόνων. Έχει διαπιστωθεί σε μελέτες πως ένας νοσηλευτής με εμπειρία στην ενδοσκόπηση μπορεί να είναι σε θέση να συντομεύσει τον χρόνο που χρειάζεται ο ιατρός για να διαβάσει τα αποτελέσματα της κάψουλας [10]. Το πόρισμα αυτό έχει επιβεβαιωθεί και από άλλες μελέτες, επισημαίνοντας πως ένας νοσηλευτής είναι απόλυτα ικανός στο να διαβάζει και στη συνέχεια να επισημαίνει παρεκκλίνουσες εικόνες [12].

Μια άλλη δυνατότητα για να μειωθεί ο χρόνος ανάγνωσης είναι η χρήση ειδικού λογισμικού για την επιλογή εικόνων στις οποίες παρουσιάζεται κάποια ανωμαλία, οι οποίες στη συνέχεια μπορούν να αναθεωρηθούν. Το λογισμικό Quickview της Given<sup>®</sup> Imaging αποτελεί παράδειγμα ενός τέτοιου λογισμικού. Η λειτουργία του συνίσταται στο να δείχνει περιοχές ενδιαφέροντος για ανασκόπηση. Το γεγονός, όμως, ότι πολύ συχνά παραλείπει εμφανείς παθολογίες καθιστά τη χρησιμότητά του μάλλον αμφίβολη [13]. Εντούτοις, σε ιδιαίτερες κλινικές καταστάσεις, όπως είναι η αδιευκρίνιστη γαστρεντερική αιμορραγία, έχει διαπιστωθεί ότι το Quickview θα μπορούσε να χρησιμοποιηθεί με αξιοπιστία, χωρίς κλινικές συνέπειες [14].

Για την ενίσχυση της απόδοσης της ενδοσκοπικής κάψουλας έχουν αναπτυχθεί επιπλέον τεχνικές, όπως είναι η εικονική χρωμοενδοσκόπηση, με τη χρήση έγχρωμων φίλτρων στις εικόνες. Η Fuji Intelligent Color Enhancement (FICE) θεωρείται ανώτερη στην ανίχνευση παθολογιών του λεπτού εντέρου, κυρίως αγγειοδυσπλασιών, σε σύγκριση με το συμβατική μέθοδο [15]. Σε μια άλλη μελέτη, η Blue Mode τεχνική βρέθηκε ανώτερη από το FICE για τη διάγνωση παθολογιών του λεπτού εντέρου [16].

Μια άλλη δυνατότητα του λογισμικού είναι η εμφάνιση, στην επιφάνεια της οθόνης του υπολογιστή, της πορείας της κάψουλας σε ένα εικονικό λεπτό έντερο, σαν ένα μετακινούμενο σημείο σε εικόνα δύο διαστάσεων [8]. Το σημείο αυτό αντιστοιχεί και στη θέση που βρίσκεται ανά πάσα στιγμή η κάψουλα. Η κλινική σημασία του αδρού αυτού προσδιορισμού της θέσης της κάψουλας έγκειται στη δυνατότητα εντοπισμού της ανατομικής θέσης της βλάβης, κατά μήκος του γαστρεντερικού σωλήνα, και συνεπώς στο σχεδιασμό της ενδεδειγμένης θεραπείας.

#### **1.2.4 Ενδείξεις**

##### **Απόλυτες**

- Απώλεια αίματος από το γαστρεντερικό σωλήνα, αδιευκρίνιστης αιτιολογίας.
- Νόσος Crohn (διάγνωση, εκτίμηση έκτασης και βαρύτητας).
- Εντεροπάθεια και έλκη από μη στεροειδή αντιφλεγμονώδη φάρμακα.
- Όγκοι του λεπτού εντέρου.

## **Σχετικές**

- Κοιλιοκάκη (διάγνωση, επιπλοκές).
- Διάγνωση απόρριψης ύστερα από μεταμόσχευση λεπτού εντέρου.
- Σύνδρομο πολυποδίασης.
- Σύνδρομο Lynch.

### **1.2.5 Αντενδείξεις**

#### **Απόλυτες**

- Ύπαρξη στενώσεων στο γαστρεντερικό σωλήνα, λόγω του κινδύνου παγίδευσης της κάψουλας και συνακόλουθης ανάγκης για χειρουργική αφαίρεση.
- Κύηση.

#### **Σχετικές**

- Βηματοδότης ή απινιδωτής.
- Διαταραχές γαστρεντερικής κινητικότητας.
- Εκκόλπωμα Zenker στον υποφάρυγγα, λόγω της αδυναμίας του ασθενούς να καταπιεί την κάψουλα. Υπάρχει δυνατότητα ενδοσκοπικής τοποθέτησής της στο δωδεκαδάκτυλο.

### **1.2.6 Δομικά μέρη**

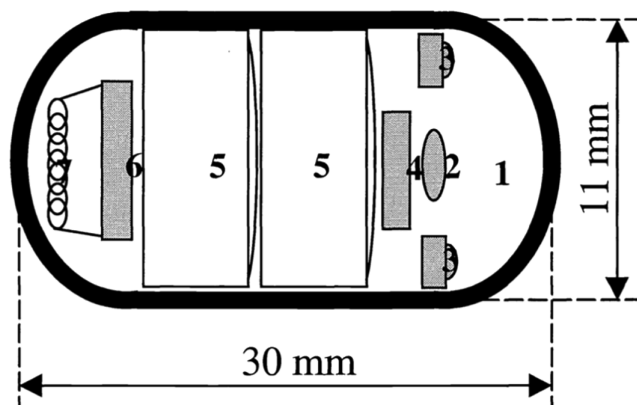
Ο σχεδιασμός της βιντεοκάψουλας έγινε εφικτός μέσω της εξέλιξης του σχεδιασμού και της απόδοσης τριών τεχνολογιών: της τεχνολογίας συμπληρωματικού ημιαγωγού μετάλλου-οξειδίου (CMOS<sup>3</sup>), των συσκευών ολοκληρωμένων κυκλωμάτων ειδικών εφαρμογών (ASIC), και των διόδων εκπομπής λευκού φωτός (LED<sup>4</sup>). Επιπλέον, ο συνδυασμός καινοτόμων σχεδίων οπτικών συστημάτων και η καλύτερη διαχείριση ενέργειας ήταν και αυτά εξίσου καθοριστικά για την ανάπτυξη της κάψουλας.

Η προσθήκη ενός κυκλώματος ενίσχυσης εξόδου σε κάθε pixel ξεχωριστά ήταν αυτό που συντέλεσε στην ραγδαία βελτίωση της εικόνας των CMOS αισθητήρων, οι οποίοι αρχικά παρείχαν έξοδο με πολύ θόρυβο. Έτσι, η ποιότητα της εικόνας μπορεί πλέον να συγκριθεί με εκείνη των CCD αισθητήρων, καταναλώνοντας όμως πολύ λιγότερη ενέργεια.

Η πρόοδος στη σχεδίαση ASIC επέτρεψε την ενσωμάτωση ενός πολύ μικρού πομπού βίντεο με αρκετή ισχύ στην έξοδό του, μεγάλη αποδοτικότητα και εύρος ζώνης (bandwidth). Η σύγχρονη εναλλαγή της λειτουργίας των LED, του CMOS, και του ASIC ελαχιστοποίησε την κατανάλωση ενέργειας.

<sup>3</sup>Complementary Metal-Oxide Semiconductor

<sup>4</sup>Light Emitting Diode



Σχήμα 1.2: Block διάγραμμα ενδοσκοπικής κάψουλας. 1: οπτικός θόλος, 2: φακός, 3: πηγή φωτός, 4: CMOS αισθητήρας, 5: μπαταρίες, 6: ASIC πομπός, 7: κεραία.

### Η CMOS κάμερα

Σε κάθε ψηφιακή κάμερα σήμερα χρησιμοποιείται μία από δύο διαφορετικές τεχνολογίες: η CCD (charge coupled device) και η CMOS. Η βασική τους λειτουργία συνίσταται στη χρήση συστοιχιών αισθητήρων οι οποίοι μετατρέπουν φωτόνια σε μετρήσιμο ηλεκτρικό ρεύμα, αναλογικά με την ένταση του φωτός. Για την επιλογή μεταξύ των δύο πρέπει να ληφθούν υπόψιν χαρακτηριστικά όπως το κόστος, το μέγεθος, η ενεργειακή κατανάλωση και η ποιότητα της εικόνας. Στα πρώτα τρία χαρακτηριστικά επικρατεί η CMOS τεχνολογία, όμως στο τελευταίο, μολονότι η ποιότητα της εικόνας στη CMOS τεχνολογία είναι επαρκώς καλή για τις περισσότερες εφαρμογές, η CCD φέρεται να την ξεπερνά [17]. Στην εργασία αυτή τα αρχεία βίντεο τα οποία αναλύονται προέρχονται από την PillCam της Given Imaging, η οποία χρησιμοποιεί αισθητήρα CMOS. Αντιθέτως, η ανταγωνιστική κάψουλα EndoCapsule της Olympus Optical χρησιμοποιεί CCD κάμερα.

Το πρώτο στάδιο για την λειτουργία της CMOS κάμερας είναι η εστίαση του φωτός, μέσω μικροσκοπικών φακών, επάνω στον αισθητήρα. Ο αισθητήρας στην πραγματικότητα αποτελείται από επιμέρους αισθητήρες, στους οποίους μπορούμε να αναφερθούμε ξεχωριστά ονομάζοντάς τους pixels. Το κάθε pixel αποτελείται από την φωτοδίοδο, η οποία μετατρέπει προσπίπτον φως σε ηλεκτρικό σήμα με τη μορφή ηλεκτρονίων ή οπών, τον διακόπτη επαναφοράς της διόδου, και τα κυκλώματα ενίσχυσης και εξόδου. Pixels με τη δυνατότητα ενίσχυσης ονομάζονται active pixels, και αποτελούν το βασικό στοιχείο των CMOS αισθητήρων.

Όταν ροή φωτονίων με ενέργεια  $E_{photon}$  προσπίπτει σε έναν ημιαγωγό με ενεργειακό διάκενο  $E_g$  τέτοιο ώστε:

$$E_g \leq E_{photon} = h \cdot \nu \quad (1.1)$$

όπου  $h$  η σταθερά του Planck και  $\nu$  η συχνότητα του φωτός, τότε ο αριθμός των απορροφώμενων φωτονίων είναι ανάλογος της έντασης της ροής τους. Επομένως, ο προσδιορισμός της φωτεινότητας για κάθε σημείο της εικόνας συνίσταται στην απλή

παρατήρηση του μεγέθους της εξόδου του αντίστοιχου pixel. Το σήμα όμως αυτό δεν μας δίνει πληροφορία για το χρώμα. Για να λάβουμε τη χρωματική απόχρωση σε συνδυασμό με την φωτεινότητα για ένα σημείο, μετρούμε την έξοδο των pixels για τρία διαφορετικά μήκη κύματος, το κόκκινο (R,  $\approx 700$  nm), το πράσινο (G,  $\approx 550$  nm) και το μπλε (B,  $\approx 440$  nm). Οι τρεις αυτές χρωματικές μεταβλητές συνθέτουν τον χρωματικό χώρο RGB, και είναι ικανές να αναπαραστήσουν μία μεγάλη γκάμμα χρωμάτων ως γραμμικό συνδυασμό τους. Επομένως, μετρώντας για κάθε περιοχή της εικόνας την ένταση του φωτός στα τρία αυτά μήκη κύματος και κβαντίζοντας το αποτέλεσμα με 8 bits, λαμβάνουμε το τελικό αρχείο εικόνας, μεγέθους  $n \times m \times 3$  (όπου  $n$  και  $m$  το πλήθος των οριζόντια και κάθετα διατεταγμένων pixels αντίστοιχα, για κάθε χρώμα), με κάθε στοιχείο του αρχείου να κυμαίνεται μεταξύ των τιμών 0 και 255.

### **ASIC πομπός**

Ένα ολοκληρωμένο κύκλωμα ειδικής εφαρμογής ASIC μπορεί να οριστεί με την ευρεία έννοια ως ένα ολοκληρωμένο κύκλωμα σχεδιασμένο για μια συγκεκριμένη εφαρμογή, όπως είναι, για παράδειγμα, ένα σύστημα τηλεπικοινωνιών. Διαφέρει από τα κλασικά ολοκληρωμένα κυκλώματα, όπως είναι οι μνήμες ή οι μικροεπεξεργαστές, στο γεγονός ότι οι τελευταίοι είναι συνήθως σχεδιασμένοι για τη χρήση τους σε μία ευρεία γκάμμα εφαρμογών.

Ο πομπός στην ενδοσκοπική κάψουλα είναι ένα ASIC σχεδιασμένο να πληρεί τις διεθνείς προδιαγραφές επικοινωνιών (όπως είναι η FCC) και χρησιμοποιείται για την μετάδοση ψηφιακών σημάτων μέσω ραδιοσυχνοτήτων με χρήση της ενσωματωμένης κεραίας. Ο πομπός επίσης χρησιμοποιείται για τον έλεγχο της φωτεινής πηγής και του CMOS αισθητήρα, με σκοπό την απενεργοποίησή τους, εφόσον αυτή είναι επιθυμητή, για την εξοικονόμηση ενέργειας της μπαταρίας. Για παράδειγμα, σε μια συσκευή κατασκευασμένη για τη καταγραφή του λεπτού εντέρου, μπορεί να είναι επιθυμητό ο πομπός να έχει ενσωματωμένη μια καθυστέρηση 2 ωρών, κατά τη διάρκεια της οποίας ο αισθητήρας εικόνας και άλλα ηλεκτρονικά στοιχεία της συσκευής είναι απενεργοποιημένα, καθώς 2 ώρες είναι κατά προσέγγιση ο χρόνος που χρειάζεται η κάψουλα για να περάσει το στομάχι και να εισέλθει στο λεπτό έντερο.

Ο κύριος σκοπός του πομπού, ωστόσο, είναι η μετάδοση των εικόνων που λαμβάνονται από τον CMOS αισθητήρα στους αισθητήρες οι οποίοι είναι επικολλημένοι στη κοιλιακής χώρα του ασθενούς [18]. Η ισχύς του σήματος μπορεί επίσης να χρησιμοποιηθεί για τον υπολογισμό της θέσης της κάψουλας στο σώμα. Οι εικόνες οι οποίες μεταδίδονται αποθηκεύονται στο φορητό μηχάνημα εγγραφής το οποίο προσαρτάται στη ζώνη.



## Κεφάλαιο 2

# Αγγειοδυσπλασίες

Η αδιευκρίνιστη γαστρεντερική αιμορραγία (ΑΓΑ) ορίζεται ως η αιμορραγία από τον γαστρεντερικό σωλήνα η οποία επιμένει ή υποτροπιάζει χωρίς ανεύρεση του αιτίου έπειτα από οισοφαγογαστροδωδεκαδακτυλοσκόπηση (EGD), κολonosκόπηση, και ακτινολογική διερεύνηση [19], [20]. Αυτή μπορεί να οφείλεται σε μικροδιαβρώσεις του οισοφάγου, στομάχου και κόλου οι οποίες δεν παρατηρήθηκαν κατά τη διάρκεια της εξέτασης, ή σε διαβρώσεις του λεπτού εντέρου οι οποίες είναι δύσκολο να παρατηρηθούν με τη συμβατική ενδοσκόπηση.

Η ΑΓΑ αποτελεί το 5% των αιμορραγιών του γαστρεντερικού και στην πλειονότητα των περιπτώσεων η εστία της αιμορραγίας εντοπίζεται στο λεπτό έντερο. Τα αίτια της αιμορραγίας του λεπτού εντέρου εξαρτώνται από την ηλικία. Σε νέους ασθενείς, ηλικίας μικρότερης των 40 ετών, τα συχνότερα αίτια αιμορραγίας είναι: όγκοι του λεπτού εντέρου (λεμφώματα, πολύποδες, καρκινοειδή, αδενοκαρκίνωμα), Μεκέλλειος απόφυση, νόσος του Crohn. Αντιθέτως, σε ασθενείς ηλικίας μεγαλύτερης των 40 ετών, οι αγγειοδυσπλασίες ευθύνονται για 30-60% των αιμορραγιών [20].

### 2.1 Παθοφυσιολογία

Οι αγγειοδυσπλασίες, πλέον γνωστές και ως αγγειεκτασίες, είναι αγγειακές ανωμαλίες του γαστρεντερικού σωλήνα. Χαρακτηριστικό των αγγειοδυσπλασιών είναι η εκτατική (διατεταμένη) μορφή των αγγείων, το λεπτό τοίχωμά τους και η απουσία ή ελάχιστη παρουσία λείων μυϊκών ινών [21]. Αν και η τυχαία εύρεσή τους χωρίς να έχουν παρουσιάσει συμπτώματα, κατά τη διάρκεια ενδοσκοπικών επεμβάσεων, είναι πολύ συχνή, αποτελούν ένα από τα κυριότερα αίτια γαστρεντερικής αιμορραγίας, ιδίως σε ηλικιωμένους. Τα τελευταία χρόνια η συχνότητα διάγνωσής τους διαρκώς αυξάνεται, πιθανότατα λόγω της βελτίωσης των μεθόδων απεικόνισης του εντέρου. Παρουσιάζονται ως περιοχές με έντονα κόκκινη απόχρωση η οποία τις διαφοροποιεί και τις κάνει να ξεχωρίζουν από τον υπόλοιπο, φυσιολογικό και διαφορετικής απόχρωσης, βλεννογόνο του λεπτού εντέρου. Το μέγεθός τους κυμαίνεται μεταξύ 2 και 10mm και συνήθως προέρχονται από ένα κεντρικό αγγείο [22].



Σχήμα 2.1: Χαρακτηριστική εικόνα αγγειοδυσπλασίας από ενδοσκοπική κάψουλα

## 2.2 Παθογένεση

Η αιτιολογία και οι μηχανισμοί για τη δημιουργία και την ανάπτυξη των αγγειοδυσπλασιών δεν είναι πλήρως κατανοητοί. Η υπόθεση με τη μεγαλύτερη αποδοχή, μέχρι στιγμής, προτάθηκε από τον Boley [23]. Η ιστολογική διερεύνηση των αγγειοδυσπλασιών έδειξε διατεταμένα και ελικοειδή αγγεία στον υποβλεννογόνο ακόμη και χωρίς προφανείς ανωμαλίες στο βλεννογόνο. Προτάθηκε ότι αυτές οι ανωμαλίες αναπτύσσονται με την ηλικία του ασθενούς λόγω χρόνιας απόφραξης των υποβλεννογόνιων φλεβών ως αποτέλεσμα της αυξημένης συσταλτικότητας στο επίπεδο των μυϊκών σιβάδων. Αυτό οδηγεί σε συμφόρηση των τριχοειδών και ανεπάρκεια των προ-τριχοειδικών σφικτήρων, που έχει ως αποτέλεσμα το σχηματισμό μικρών αρτηριοφλεβικών επικοινωνιών [21].

Αυτή η θεωρία υποστηρίζεται μερικώς από το γεγονός ότι οι αγγειοδυσπλασίες εμφανίζονται συχνά σε γηραιότερους ασθενείς. Καθώς το έντερο έχει κυλινδρική μορφή, ισχύει ο νόμος Laplace:

$$T = \frac{Pr}{2t} \quad (2.1)$$

όπου  $T$  η τάση (δύναμη ανά μονάδα μήκους) στα τοιχώματα,  $P$  η πίεση,  $r$  η ακτίνα και  $t$  το πάχος των τοιχωμάτων. Από την εξίσωση φαίνεται πως η τάση είναι ανάλογη της ακτίνας και αντιστρόφως ανάλογη του πάχους των τοιχωμάτων. Επομένως, αυτός ο μηχανισμός μπορεί να εξηγήσει την ανάπτυξη εκτατικών αγγείων και αγγειοδυσπλασιών σε σημεία όπου η τάση στα τοιχώματα είναι αρκετά μεγάλη ώστε να συμπιέζονται φλεβίδια αλλά όχι αρτηρίδια. Τέτοια σημεία υπάρχουν στο ανιόν κόλον και τυφλό, όπου η διάμετρος είναι μεν μεγάλη, αλλά τα τοιχώματα είναι λεπτά. Παρόμοιος μηχανισμός συναντάται και στον πυλωρό και τη γαστρική κοιλότητα όπου η συστολή των μυϊκών σιβάδων είναι ιδιαίτερα έντονη.

Μερικοί ερευνητές έχουν παρατηρήσει ότι ασθενείς με αγγειοδυσπλασίες είναι

πιο πιθανό να εμφανίζουν καρδιακές, αγγειακές ή πνευμονικές ασθένειες, καταλήγοντας στο συμπέρασμα ότι η βλεννογονική ισχαιμία από χρόνια υποξία ή υποαιμάτωση μπορεί να συμβάλει στην ανάπτυξη αγγειοδυσπλασιών. Είναι σημαντικό, όμως, να αναφερθεί πως αυτές οι έρευνες ήταν μικρές μελέτες για τον έλεγχο των περιπτώσεων αγγειοδυσπλασιών χωρίς ιστολογική συσχέτιση. Αυτό οδήγησε στην πρόταση ότι η αυξημένη παρατήρηση αιμορραγουσών αγγειοδυσπλασιών σε ασθενείς με αορτική στένωση μπορεί να οφείλεται σε μειωμένη καρδιακή παροχή και ιστική αιμάτωση.

## **2.3 Διάγνωση**

Η έκταση του εύρους των εξετάσεων στις οποίες θα πρέπει να υποβληθεί ένας ασθενής με αδιευκρίνιστη γαστρεντερική αιμορραγία εξαρτάται σε μεγάλο βαθμό από τη σοβαρότητα της αιμορραγίας και της κλινικής περίπτωσης. Οι κύριες μέθοδοι για τη διάγνωση των αγγειοδυσπλασιών είναι:

### **Ενδοσκοπική απεικόνιση**

Η ενδοσκόπηση αποτελεί το κύριο εργαλείο για τη διάγνωση των αγγειοδυσπλασιών, κυρίως λόγω των πολύ σημαντικών βελτιώσεων στον εξοπλισμό και την ανάλυση της εικόνας. Οι αγγειοδυσπλασίες στο ανώτερο πεπτικό και στο κόλον συνήθως διαγιγνώσκονται με χρήση της τυπικής γαστροσκόπησης και κολονοσκόπησης αντίστοιχα. Ειδικά για το λεπτό έντερο όμως, στα διαγνωστικά εργαλεία συγκαταλέγονται το εντεροσκόπιο και η ενδοσκοπική κάψουλα.

Ανωμαλίες όπως είναι οι αγγειοδυσπλασίες μπορεί πολύ συχνά να μην γίνουν αντιληπτές κατά τη διάρκεια της ενδοσκόπησης για μια πληθώρα λόγων, μεταξύ των οποίων είναι και η εμπειρία του χειριστή του ενδοσκοπίου, η χαμηλή ορατότητα καθώς και το μέγεθος αλλά και η θέση της αγγειοδυσπλασίας, ιδιαίτερος εάν βρίσκεται πίσω από πτυχές του βλεννογόνου [24]. Είναι επίσης πιθανό οι αγγειοδυσπλασίες να διαγνωστούν λανθασμένα ως μια τοπική περιοχή φλεγμονής ή τραύματος [25]. Ως συνέπεια αυτού προτείνεται η εκτέλεση πολλαπλών εξετάσεων, ιδίως σε περιπτώσεις υψηλής κλινικής υποψίας πριν την εξέταση του λεπτού εντέρου.

### **Ακτινογραφική απεικόνιση**

Οι ακτινογραφικές τεχνικές απεικόνισης χρησιμοποιούνται μόνο σε ασθενείς οι οποίοι εμφανίζουν σημεία ενεργούς ΑΓΑ είτε για τη διάγνωση, είτε για την επιβεβαίωση της τοποθεσίας της αιμορραγικής εστίας. Οι εξετάσεις αυτές γίνονται συνήθως με χρήση αξονικής αγγειογραφίας (CTA<sup>1</sup>), μαγνητικής αγγειογραφίας ή κλασικής αγγειογραφίας.

---

<sup>1</sup>Computed Tomography Angiography

## 2.4 Θεραπεία

Όσον αφορά τη θεραπεία των αγγειοδυσπλασιών, δεν υπάρχει ακόμη ομόφωνη απάντηση για το ποιά είναι η καλύτερη μέθοδος. Σε περίπτωση παρουσίας αγγειοδυσπλασιών οι οποίες δεν αιμορραγούν, η θεραπεία εξαρτάται από το γενικότερο πλαίσιο της κλινικής περίπτωσης. Για παράδειγμα, δεν προτείνεται θεραπεία για αγγειοδυσπλασίες οι οποίες ανακαλύπτονται τυχαία σε ασθενείς με συμπτώματα τα οποία δεν σχετίζονται με αιμορραγία. Τέτοιοι ασθενείς συνήθως παραμένουν ασυμπτωματικοί και στο μέλλον, καθώς η πιθανότητα μελλοντικής αιμορραγίας σε αυτές τις περιπτώσεις είναι χαμηλή [26]. Επίσης, ασθενείς οι οποίοι παρουσιάζουν αγγειοδυσπλασίες σε τουλάχιστον ένα μέρος του γαστρεντερικού σωλήνα συνήθως παρουσιάζουν επιπλέον αγγειοδυσπλασίες και αλλού [27]. Σε περίπτωση ασθενών με ΑΓΑ, η θεραπεία των αγγειοδυσπλασιών εξαρτάται από το μέγεθος, τη θέση και τον αριθμό τους, όπως επίσης και από την κλινική σοβαρότητα της αναιμίας ή απώλειας αίματος.

Αν και οι μέθοδοι με τις οποίες γίνεται η θεραπεία των αγγειοδυσπλασιών είναι πολλές, θα παρατεθούν μόνο μερικές από αυτές, επιλεγμένες λόγω της σπουδαιότητάς τους ή της συχνής χρήσης τους.

### Θερμοπηξία με Argon Plasma

Η θερμοπηξία με Argon Plasma<sup>2</sup> (APC<sup>3</sup>) είναι η παραδοσιακή μέθοδος για τη θεραπεία αγγειοδυσπλασιών κατά τη διάρκεια κολonosκόπησης, ενδοσκόπησης ανωτέρου πεπτικού και εντεροσκόπησης [28]. Πρόκειται για μία ασφαλή θεραπευτική διαδικασία με πολύ λίγες παρενέργειες [29]. Η μέθοδος αυτή χρησιμοποιεί συνδυασμό ηλεκτρικού ρεύματος και αερίου αργού. Εφαρμόζοντας μια μεγάλη τάση της τάξης των 6kV στο αργό, αυτό ιονίζεται επιτρέποντας τη μεταφορά ρεύματος υψηλής συχνότητας χωρίς άμεση επαφή στην επιθυμητή ανωμαλία ή ιστό. Ως αποτέλεσμα αυτού έχουμε θερμοπηξία της αιμορραγικής εστίας. Το γεγονός ότι δεν έχουμε άμεση επαφή με τον ιστό είναι αυτό που καθιστά την όλη διαδικασία ασφαλή έπειτα από καθαρισμό του εντέρου από αέρια και επιτρέπει τη χρήση του σε σημεία του γαστρεντερικού σωλήνα με λεπτά τοιχώματα, όπως είναι το τυφλό.

### Αιμοστατικά κλιπ

Μια άλλη αποτελεσματική και ασφαλής μέθοδος είναι η χρήση αιμοστατικών κλιπ κατά τη διάρκεια ενδοσκόπησης [30], [31]. Τα κλιπ είναι ιδιαίτερα χρήσιμα σε περιπτώσεις μεμονωμένων και ιδιαίτεως μεγάλων σε μέγεθος ανωμαλιών, όπου με την εφαρμογή τους αποκόπουν το κύριο αγγείο το οποίο και τροφοδοτεί την αγγειοδυσπλασία. Μπορεί να χρησιμοποιηθεί σε συνδυασμό με τη θερμοπηξία με σκοπό την αποφυγή περαιτέρω αιμορραγικών επιπλοκών.

---

<sup>2</sup>Πλάσμα αργού

<sup>3</sup>Argon Plasma Coagulation

## **Χειρουργείο**

Λόγω της προόδου στις ενδοσκοπικές και αγγειογραφικές διαγνωστικές και θεραπευτικές τεχνικές η ανάγκη για χειρουργική επέμβαση σε ασθενείς με αγγειοδυσπλασίες έχει πλέον περιοριστεί. Παρόλα αυτά, η χειρουργική εκτομή πραγματοποιείται ακόμη σε ασθενείς με σοβαρή αιμορραγία η οποία είναι αδύνατον να ελεγχθεί ή περιοριστεί με άλλες μεθόδους, ή σε ασθενείς με χρόνια αιμορραγία οι οποίοι εξαρτώνται από μετάγγιση αίματος όταν ο εντοπισμός της αιμορραγικής εστίας είναι σαφής.

## **Φαρμακευτικές θεραπείες**

Η θεραπεία των αγγειοδυσπλασιών με χρήση αγγειογραφικών, ενδοσκοπικών και χειρουργικών τεχνικών μπορεί να έχει σοβαρές επιπλοκές, και σε ορισμένες περιπτώσεις μπορεί να είναι ανεπαρκής στη πρόληψη νέας αιμορραγίας στο μέλλον [21]. Αυτό μπορεί να οφείλεται στο γεγονός ότι είναι δύσκολο να βρεθούν όλες οι ακριβείς θέσεις των ανωμαλιών αυτών, ιδιαίτερα στο λεπτό έντερο [32]. Επομένως, μια εναλλακτική μέθοδος είναι η χορήγηση φαρμάκων ή ορμονών (οιστρογόνο και προγεστερόνη).

## **Άλλες θεραπείες**

Άλλες θεραπείες οι οποίες χρησιμοποιούνται, είναι η σκληροθεραπεία [33], η φωτοπηξία [34], η ηλεκτροπηξία [35] (η οποία ήταν και η πρώτη ενδοσκοπική τεχνική για τη θεραπεία αγγειοδυσπλασιών) και η διαδερμική αγγειογραφική παρέμβαση (ΤΑΙ<sup>4</sup>).

---

<sup>4</sup>Transcatheter Angiography and Intervention



## Κεφάλαιο 3

# Ανάλυση εικόνας

Στην όραση υπολογιστών ο τελικός στόχος είναι η περιγραφή του κόσμου όπως τον βλέπουμε μέσα από μία ή περισσότερες εικόνες και η ανακατασκευή των ιδιοτήτων του, όπως είναι το σχήμα, η φωτεινότητα και η κατανομή του χρώματος. Η λύση τέτοιων και παρόμοιων προβλημάτων αποτελεί ένα εξαιρετικά δύσκολο εγχείρημα του οποίου η πολυπλοκότητα συχνά υποτιμάται. Είναι πολύ γνωστή η ιστορία σύμφωνα με την οποία ο Marvin Minsky, καθηγητής στο MIT, ζήτησε το 1966 από έναν προπτυχιακό φοιτητή με το όνομα Gerald Jay Sussman να συνδέσει μια κάμερα σε έναν υπολογιστή, και κατά τη διάρκεια του καλοκαιριού να κατασκευάσει ένα πρόγραμμα το οποίο θα περιγράφει την εικόνα που βλέπει η κάμερα [37]. Όπως αποδείχτηκε αργότερα, το εγχείρημα αυτό ήταν πολύ πιο δύσκολο από ότι είχε εκτιμηθεί στην αρχή, και έδωσε το έναυσμα για τη δημιουργία ενός ολόκληρου κλάδου, αυτού της όρασης υπολογιστών. Είναι πραγματικά εκπληκτικό το πώς κάτι που ο άνθρωπος και τα ζώα μπορούν να κάνουν χωρίς ιδιαίτερη προσπάθεια είναι πολύ δύσκολο να υλοποιηθεί μέσω αλγορίθμων και να επιτευχθεί από έναν υπολογιστή. Πλέον, έπειτα από 50+ χρόνια εξέλιξης, η όραση υπολογιστών χρησιμοποιείται ευρέως για μία μεγάλη γκάμα εφαρμογών η οποία συμπεριλαμβάνει:

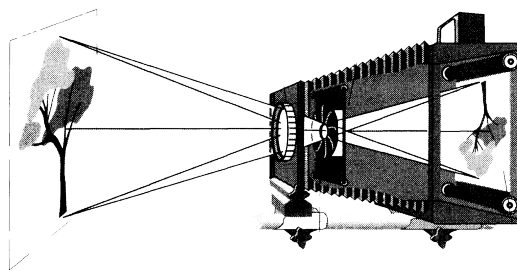
- Αναγνώριση οπτικών χαρακτήρων (OCR): Η ανάγνωση γραμμάτων και αριθμών, χειρόγραφων ή μη, όπως είναι για παράδειγμα ο ταχυδρομικός κώδικας στην αλληλογραφία, ή η αναγνώριση του αριθμού κυκλοφορίας στις πινακίδες των αυτοκινήτων.
- Μηχανική επιθεώρηση: οπτική επιθεώρηση με τη χρήση στερεοσκοπικής όρασης και ειδικού φωτισμού για την εξακρίβωση της ποιότητας κατασκευών όπως είναι τα πτερά ενός αεροπλάνου ή τα εξαρτήματα αυτοκινήτων.
- Εμπόριο: Αυτόματη αναγνώριση αντικειμένων για αυτόματους πωλητές σε καταστήματα.
- 3D ανακατασκευή: Αυτόματη ανακατασκευή του τρισδιάστατου μοντέλου ενός αντικειμένου, για παράδειγμα ανακατασκευή κτιρίων από εναέριες φωτογραφίες και χρήση τους σε συστήματα όπως είναι οι χάρτες.

- Ιατρική απεικόνιση: Αυτόματη αναγνώριση παθολογιών, μελέτη διαφορών σε προ-εγχειρητικές και μετα-εγχειρητικές εικόνες ή μακροχρόνιες μελέτες για τη μεταβολή οργάνων, όπως είναι ο εγκέφαλος, στον άνθρωπο καθώς αυτός μεγαλώνει.

### 3.1 Αρχεία εικόνας

Όλοι μας είμαστε εξοικειωμένοι με τις φωτογραφίες, την απεικόνιση, ή αλλιώς την προβολή, του τρισδιάστατου κόσμου μας στις δύο διαστάσεις. Η κατανόηση του σχηματισμού των εικόνων είναι απαραίτητη για τη χρησιμοποίησή τους προς εξαγωγή πληροφορίας για τον κόσμο.

Η βασική αρχή κάθε φωτογραφικής μηχανής έγκειται στη συλλογή ακτίνων φωτός οι οποίες, περνώντας μέσα από το διάφραγμα και ενός συστήματος φακών, προσπίπτουν επάνω σε έναν αισθητήρα εικόνας ο οποίος μπορεί να είναι είτε αναλογικό φιλμ, στην περίπτωση της κλασικής (αναλογικής) φωτογραφίας, είτε ένα image sensing chip στην περίπτωση της ψηφιακής φωτογραφίας.



Σχήμα 3.1: Τμήματα κάμερας με φακό και σχηματισμός εικόνας (Από Hecht [38])

Ο παραδοσιακός τύπος φωτογραφίας περιλαμβάνει την αποτύπωση της εικόνας επάνω σε ένα φιλμ ως αναλογικό, μόνο, σήμα. Λέγοντας ότι ένα σήμα (σε αυτή την περίπτωση η εικόνα μας) είναι αναλογικό, εννοούμε πως πρόκειται για ένα συνεχές στο χωροχρόνο σήμα, καθώς και πως αυτό λαμβάνει τιμές μέσα από ένα συνεχές σύνολο αριθμών. Ο σημερινός τρόπος αποθήκευσης και επεξεργασίας φωτογραφιών, όμως, επιβάλλει τη χρήση ψηφιακών μέσων και συνεπώς σημάτων. Σε απόλυτη αντιστοιχία με το αναλογικό σήμα, με τον όρο ψηφιακό σήμα εννοούμε ένα σήμα το οποίο είναι διακριτό στο χώρο και στο χρόνο, και λαμβάνει τιμές μέσα από ένα διακριτό σύνολο τιμών. Πριν από οποιαδήποτε ψηφιακή επεξεργασία ή αποθήκευση, λοιπόν, ενός σήματος, είναι απαραίτητη η μετατροπή του από αναλογικό σε ψηφιακό (μέθοδος γνωστή ως A/D conversion). Η μετατροπή αυτή αποτελείται από δύο διαδικασίες, τη δειγματοληψία και την κβάντιση.



### 3.1.1 Δειγματοληψία

Η δειγματοληψία είναι η διαδικασία με την οποία μετατρέπουμε ένα συνεχούς χώρου σήμα σε ένα σήμα διακριτού χώρου. Ένα δισδιάστατο σήμα εικόνας, δηλαδή, κατατάσσεται σε έναν δισδιάστατο πίνακα μεγέθους  $n \times m$  με τη μορφή αριθμών, οι οποίοι είναι ανάλογοι της φωτεινότητας κάθε σημείου. Κάθε στοιχείο αυτού του δισδιάστατου πίνακα αποκαλείται *picture element*, δηλαδή στοιχείο εικόνας, και για συντομία καλείται με τη γνωστή ονομασία *pixel*. Συνδυάζοντας τα *pixels* μπορούμε να συνθέσουμε, κατά προσέγγιση, την αρχική εικόνα. Όσο μεγαλύτερη είναι η δειγματοληψία τόσο περισσότερα τα *pixels*, δηλαδή τόσο μεγαλύτερο το μέγεθος του πίνακα, και, συνεπώς, τόσο πιο λεπτομερής θα είναι και προκύπτουσα εικόνα. Μαθηματικά, εάν η μέγιστη τιμή συχνότητας που περιέχεται σε ένα αναλογικό σήμα  $x_a(t)$  είναι η  $F_{\max}$  και η δειγματοληψία του σήματος γίνεται με τιμή συχνότητας  $F_s \geq 2F_{\max}$ , τότε το σήμα  $x_a(t)$  μπορεί να ανακατασκευαστεί πλήρως από τις τιμές που έχουν καταγραφεί.

Ο αριθμός των στηλών και γραμμών του πίνακα συχνά επιλέγεται ώστε να είναι δύναμη του 2, διότι αυτό απλουστεύει την διευθυνσιοδότηση των δειγμάτων από τον υπολογιστή, καθώς επίσης και επειδή ορισμένοι αλγόριθμοι, όπως ο διακριτός μετασχηματισμός Fourier, είναι ιδιαίτερα αποδοτικοί όταν εφαρμόζονται σε σήματα με διαστάσεις δυνάμεις του 2.

### 3.1.2 Κβάντιση

Το δεύτερο κομμάτι της μετατροπής είναι η κβάντιση. Έχουμε ήδη αναφέρει πως το κάθε στοιχείο της εικόνας έχει τη μορφή ενός αριθμού. Ο αριθμός αυτός αποκαλείται συχνά και ένταση του *pixel*, καθώς είναι ανάλογος της έντασης του σήματος στον αισθητήρα, δηλαδή του πλήθους των φωτονίων που προσέπεσαν στο αντίστοιχο αισθητηριακό *pixel*. Με τη διαδικασία της κβάντισης μετατρέπουμε τους αριθμούς αυτούς από συνεχείς σε διακριτούς, και το εύρος των αριθμών αυτών από συνεχές σε διακριτό. Αυτό συνήθως επιτυγχάνεται με στρογγυλοποίηση ή κάποια άλλη, μη-αναστρέψιμη, μη-γραμμική διαδικασία καταστροφής πληροφορίας.

Όταν κβαντίζουμε τη γκριζα διαβάθμιση ή στάθμη (*grey level*) μιας εικόνας, της αναθέτουμε αριθμούς μέσα από ένα πεπερασμένο πλήθος, το οποίο αποτελεί το εύρος του γκριζου. Αφού το εύρος οριστεί, μια απλή και αποδοτική μέθοδος κβάντισης είναι η απλή στρογγυλοποίηση της έντασης του κάθε *pixel* στην κοντινότερη τιμή η οποία περιέχεται μέσα στο εύρος αυτό. Οι τιμές αυτές μπορεί να είναι οποιοσδήποτε, όμως συνήθως, για την ευκολότερη ψηφιακή αναπαράσταση, επιλέγονται μέσω ενός γραμμικού μετασχηματισμού σε ένα πεπερασμένο σύνολο θετικών ακεραίων  $(0, \dots, K - 1)$  όπου  $K$  είναι δύναμη του 2:  $K = 2^B$ . Συνεπάγεται ότι οι τα επιτρεπτά επίπεδα γκρι είναι  $K$ , και ο αριθμός των *bits* για την περιγραφή της έντασης κάθε *pixel* είναι  $B$ .

Στα συνηθισμένα *formats* ψηφιακών εικόνων (PNG, JPEG, TIFF, κ.ά) οι μο-



Σχήμα 3.2: Μονόχρωμη εικόνα κβαντισμένη με διαφορετικό πλήθος bits για κάθε pixel.

νόχρωμες εικόνες κβαντίζονται με 8 bits/pixel, διαθέτουν δηλαδή 255 επιτρεπτά επίπεδα του γκρι για κάθε εικονοστοιχείο, και οι έγχρωμες εικόνες με 24 bits/pixel λόγω των 3 καναλιών τα οποία είναι κβαντισμένα με 8 bit/pixel το καθένα. Η διαφορά της κβάντισης στις εικόνες αυτές από αυτή που περιγράφηκε πιο πάνω, έγκειται στο ότι δεν γίνεται σε γραμμική κλίμακα φωτεινότητας αλλά σε μη-γραμμική τύπου log. Ο λόγος είναι ότι το ανθρώπινο μάτι αντιλαμβάνεται την ένταση του φωτός και του χρώματος με μη-γραμμικό τρόπο, οπότε είναι προτιμότερο οι εικόνες να κωδικοποιούνται ώστε να χρησιμοποιούν περισσότερα bits σε λεπτομέρειες στις οποίες είναι πιο ευαίσθητος ο ανθρώπινος οφθαλμός. Μοντέρνες ψηφιακές κάμερες χρησιμοποιούν 12 bits/pixel για μονόχρωμες εικόνες σε RAW format.

Συνοψίζοντας, μια μονόχρωμη ψηφιακή εικόνα αποτελείται από ένα σύνολο  $n \times m$  κέραιων αριθμών  $B$  bits, με εύρος  $(0, \dots, 2^B - 1)$  οι οποίοι αποτελούν τη γκριζα διαβάθμιση της εικόνας και είναι διατεταγμένοι σε έναν  $n \times m$  πίνακα τα στοιχεία του οποίου ονομάζονται pixels, ή, αλλιώς, εικονοστοιχεία.

Όπως θα εξηγηθεί παρακάτω, με ανάλογο τρόπο ορίζονται και οι έγχρωμες εικόνες, οι οποίες όμως αποτελούνται από πολλαπλά κανάλια, ο γραμμικός συνδυασμός των οποίων συνθέτει την έγχρωμη εικόνα μας.

### 3.2 Γειτονικές πράξεις (Neighborhood operations)

Έχουμε ήδη εξηγήσει το πώς αναπαρίστανται χωρικά οι ψηφιακές εικόνες. Η ανάλυση των χωρικών σχέσεων των γκριζών τιμών σε μια μικρή περιοχή, ή αλλιώς γειτονιά, αποτελεί το πρώτο βήμα για την αναγνώριση αντικειμένων στις εικόνες. Ας θεωρήσουμε μια εικόνα η οποία περιέχει αντικείμενα ομοιόμορφης φωτεινότητας. Αν η τιμή του γκριζου μέσα σε μια μικρή γειτονιά παραμένει σχεδόν αμετάβλητη,

τότε η γειτονιά αυτή βρίσκεται μέσα σε ένα από τα εικονιζόμενα αντικείμενα. Αντιθέτως, εάν η γκριζα τιμή μεταβάλλεται απότομα, η γειτονιά βρίσκεται επάνω στο όριο (ακμή) κάποιου αντικειμένου. Με αυτόν τον τρόπο αναγνωρίζουμε περιοχές σταθερών γκριζών τιμών και ακμές.

Αν ελέγξουμε κάθε pixel μιας εικόνας όπως του παραπάνω παραδείγματος ξεχωριστά, είναι αδύνατο να φτάσουμε σε κάποιο συμπέρασμα για το αν βρίσκεται επάνω σε ένα αντικείμενο ή επάνω σε μια ακμή. Αυτό συμβαίνει γιατί δεν μας είναι αρκετή μόνο η πληροφορία για το υπό εξέταση pixel, αλλά μας ενδιαφέρει και η σύγκρισή του με τα υπόλοιπα pixels μέσα σε μια γειτονιά. Για την επίτευξη τέτοιων λειτουργιών από τον υπολογιστή είναι απαραίτητος ο ορισμός ενός νέου τύπου υπολογισμών οι οποίοι συνδυάζουν γειτονικά pixels με κατάλληλο τρόπο και δίνουν ως αποτέλεσμα μια καινούργια εικόνα. Υπολογισμοί τέτοιου τύπου ονομάζονται γειτονικοί υπολογισμοί ή *γειτονικές πράξεις* (neighborhood operations). Αποτελούν βασικά εργαλεία για την *χαμηλού-επιπέδου ανάλυση εικόνας* (low-level image processing).

Το αποτέλεσμα κάθε γειτονικής πράξης είναι μια εικόνα. Όμως, το περιεχόμενο της εικόνας αυτής έχει αλλάξει σε σχέση με το αρχικό. Για παράδειγμα, μια γειτονική πράξη η οποία έχει σχεδιαστεί έτσι ώστε να ανιχνεύει ακμές θα πρέπει να δίνει ως αποτέλεσμα μια εικόνα στην οποία τα μοναδικά pixels με υψηλή τιμή ανήκουν σε ακμές. Μια σημαντική παρατήρηση είναι ότι όταν εφαρμόζουμε μια γειτονική πράξη συνήθως χάνουμε πληροφορία. Δεν μπορούμε να εξάγουμε από την προκύπτουσα εικόνα τις τιμές της αρχικής. Για αυτόν ακριβώς τον λόγο οι γειτονικές πράξεις αποκαλούνται επίσης και *φίλτρα*, καθώς εξάγουν μόνο ένα επιθυμητό χαρακτηριστικό από μια εικόνα. Είναι βέβαιο πως ο όρος «φίλτρα» δεν είναι άγνωστος στον αναγνώστη, κυρίως διότι ζούμε σε μια εποχή όπου οι ψηφιακές εικόνες και φωτογραφίες είναι ευρύτατα διαδεδομένες. Όμως τα φίλτρα στην όραση υπολογιστών δεν είναι τα εφέ τα οποία προσθέτει κάποιος σε μία εικόνα πριν τη διαμοιραστεί με τους φίλους του στο διαδίκτυο. Αυτά συνήθως αποτελούνται από ένα σύνολο χαμηλού-επιπέδου πράξεων ανάλυσης εικόνας, όπως είναι η αλλαγή της αντίθεσης (contrast), του χρώματος, της οξύτητας, κ.ο.κ.

### 3.2.1 Γενικός ορισμός

Μια γειτονική πράξη  $N$  λαμβάνει τις τιμές των pixels τα οποία περιβάλλουν ένα σημείο, πραγματοποιεί κάποιους υπολογισμούς συνδυάζοντας τις τιμές αυτές κατάλληλα, και γράφει το αποτέλεσμα επάνω στο pixel.

Ας θεωρήσουμε μια ψηφιακή εικόνα  $f$  η οποία αποτελείται από έναν πίνακα μεγέθους  $m \times n$ . Μια *διακριτή γειτονική πράξη* δημιουργεί μια καινούργια εικόνα  $g$  μέσω της ακόλουθης πράξης:

$$g = N(f_{i-m,j-n}, \forall [i, j]^T \in \mathbb{M}) \quad (3.1)$$

όπου  $\mathbb{M}$  η *μάσκα*, το *παράθυρο*, ή το *δομικό στοιχείο* της γειτονικής πράξης. Το

πρωτεύον χαρακτηριστικό σε μια γειτονική πράξη είναι το μέγεθος της γειτονιάς. Το παράθυρο μπορεί να είναι ορθογώνιο ή οποιαδήποτε άλλη μορφή. Συνήθως το παράθυρο είναι και συμμετρικό, δηλαδή το αποτέλεσμα της πράξης αποθηκεύεται στο κεντρικό pixel του παραθύρου. Για αυτό το λόγο, το μέγεθος του παραθύρου επιλέγεται έτσι ώστε να είναι περιττός αριθμός, διαφορετικά δεν υπάρχει pixel το οποίο να βρίσκεται στο κέντρο της μάσκας.

### 3.2.2 Γραμμικό φιλτράρισμα

Τώρα, από τον πίνακα της αρχικής εικόνας  $f$  διάστασης  $m \times n$  όπως και πριν, ας κατασκευάσουμε έναν καινούργιο πίνακα  $g$ , γेमίζοντας κάθε σημείο, ή αλλιώς pixel, του πίνακα αυτού με το σταθμισμένο σύνολο των τιμών των pixels που περιβάλλουν την ανίστοιχη θέση του σημείου στην εικόνα, χρησιμοποιώντας κάθε φορά το ίδιο βάρος  $w = 1$ . Μπορούμε επίσης να διαιρέσουμε την προκύπτουσα τιμή με το πλήθος των pixels που χρησιμοποιήθηκαν, λαμβάνοντας έτσι το μέσο όρο της τιμής των γειτονικών pixel για κάθε pixel στην εικόνα μας. Δηλαδή, σε ένα τυχαίο σημείο  $(i, j)$  θα ισχύει:

$$g(i, j) = \frac{1}{(2k + 1)^2} \sum_{u=-k}^k \sum_{v=-k}^k f(i + u, j + v) \quad (3.2)$$

Η διαδικασία αυτή ονομάζεται *γραμμικό φιλτράρισμα* (linear filtering) και μόλις εφαρμόσαμε ένα φίλτρο εξομάλυνσης της εικόνας, γνωστό ως box blur.

### 3.2.3 Συνέλιξη

Μπορούμε να αλλάξουμε το αποτέλεσμα του φιλτραρίσματος αλλάζοντας τα βάρη τα οποία χρησιμοποιήσαμε. Το μοτίβο των βαρών τα οποία χρησιμοποιούνται σε ένα γραμμικό φιλτράρισμα συνήθως αποκαλείται «πυρήνας» (kernel) του φίλτρου και είναι αυτό το οποίο μέχρι πρότινος αποκαλούσαμε «μάσκα». Η διαδικασία με την οποία εφαρμόζουμε το φίλτρο ονομάζεται *συνέλιξη*. Δοθέντος ενός πυρήνα  $h$ , η συνέλιξη του πυρήνα με μια εικόνα  $f$  μας δίνει μια εικόνα  $g$  για την οποία, ομοίως με πριν, το  $(i, j)$  στοιχείο της δίνεται από:

$$g(i, j) = \sum_{u,v} f_{u,v} h_{i-u, j-v} \quad (3.3)$$

Η διαδικασία αυτή ορίζει τη συνέλιξη, και λέμε ότι συνελίξαμε το  $h$  με το  $f$  για να προκύψει το  $g$ . Τη συμβολίζουμε ως:

$$g = f * h \quad (3.4)$$

Πρέπει να δοθεί ιδιαίτερη προσοχή στο ότι η κατεύθυνση της μεταβλητής  $u$ , καθώς και της  $v$ , έχει αναστραφεί. Αυτό οφείλεται στο μαθηματικό ορισμό της συνέλιξης, ο οποίος είναι:

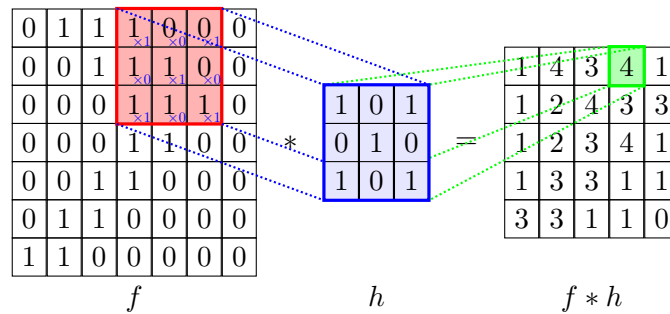
$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)g(t - \tau)d\tau \quad (3.5)$$

Η διαφορά της συνέλιξης στην όραση υπολογιστών με τον ορισμό της μαθηματικής συνέλιξης που αναφέρθηκε πιο πάνω έγκειται στο ότι η μια είναι διακριτή, και η άλλη συνεχής. Η διακριτή συνέλιξη αποτελεί μία πάρα πολύ σημαντική πράξη την οποία αξίζει να δει κανείς λεπτομερώς.

Όπως έχει ήδη παρατηρηθεί, για την πραγματοποίηση της συνέλιξης κατοπριζουμε είτε τον πυρήνα (ή αλλιώς τη μάσκα), είτε την ίδια την εικόνα. Αυτό φαίνεται από τα αρνητικά πρόσημα των  $u$  και  $v$  στην εξίσωση 3.3. Ας ορίσουμε έναν πυρήνα:

$$h = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad (3.6)$$

Είναι διάστασης  $3 \times 3$ , και το γεγονός ότι το πλήθος των στοιχείων του είναι περιττός αριθμός (9), σημαίνει ότι υπάρχει και κεντρικό pixel. Επίσης εδώ έχει οριστεί ως απόλυτα συμμετρικός, το οποίο σημαίνει ότι δεν αλλάζει ακόμη και μετά την κατόπτρισή του. Για να υπολογίσουμε το αποτέλεσμα σε ένα σημείο  $(i, j)$ , τοποθετούμε το κέντρο του πυρήνα στο σημείο αυτό, υπολογίζουμε το σταθμισμένο άθροισμα όλων των στοιχείων μέσα στη γειτονιά, και τοποθετούμε το αποτέλεσμα στο αντίστοιχο σημείο του καινούργιου πίνακα.



Σχήμα 3.3: Διακριτή δισδιάστατη συνέλιξη.

Άμεσα θα παρατηρήσει κάποιος ότι η νέα, προκύπτουσα εικόνα είναι διάστασης μικρότερης από την αρχική. Ο λόγος γίνεται εύκολα αντιληπτός αν προσπαθήσει κανείς να κάνει τον υπολογισμό κοντά στα άκρα της εικόνας. Εκεί εμφανίζεται ένα δίλημμα, αφού μέρος του πυρήνα βρίσκεται έξω από την εικόνα, και εκεί δεν υπάρχουν pixels ή αντίστοιχες τιμές τους.

Μία λύση είναι να μην υπολογίσουμε καθόλου τη συνέλιξη στα άκρα της εικόνας, όπως ακριβώς έγινε και στο σχήμα 3.3. Όμως, άμεση συνέπεια αυτού, είναι ότι η εικόνα θα μικρύνει. Μία άλλη λύση είναι να εφαρμόσουμε μια τεχνική γνωστή ως *padding*, με την οποία αυξάνουμε τεχνητά το μέγεθος της αρχικής εικόνας τόσο ώστε ο πυρήνας να μην πέφτει έξω από την εικόνα όταν βρισκόμαστε στα προηγούμενα

όρια. Τις τιμές των νέων pixels τις ορίζουμε αυθαίρετα. Υπολογίζουμε και πάλι τη συνέλιξη στη νέα, μεγαλύτερη λόγω του padding εικόνα μας, η οποία και πάλι μειώνεται σε μέγεθος όμως το νέο μέγεθος είναι ίδιο με την αρχική εικόνα πριν εφαρμόσουμε το padding. Υπάρχουν πολλές επιλογές για την εφαρμογή του padding, μερικές εκ των οποίων είναι:

- zero-padding: Όλα τα pixels έξω από την αρχική εικόνα τίθενται ίσα με το 0.
- constant: Όλα τα pixels παίρνουν την τιμή μιας σταθεράς από τα όρια της εικόνας.
- clamp: Τα pixels στα όρια της εικόνας επαναλήπτονται επ' αόριστον.
- cyclic: Εφαρμόζουμε στην ουσία μια κυκλική συνέλιξη, παίρνοντας, όταν είμαστε στα όρια, τιμές από την άλλη μεριά της εικόνας.

Πρέπει να αναφέρουμε ότι καμία μέθοδος για τη συνέλιξη δεν είναι τέλεια, όμως κάθε φορά μια από αυτές θα πρέπει να επιλεγεί ανάλογα με τις ανάγκες του προβλήματος.

Γενικά, όταν συνελίσσουμε μία εικόνα  $n \times m$  με ένα φίλτρο  $f$  μεγέθους  $f_w \times f_h$  παράγεται μία νέα εικόνα μεγέθους  $N \times M$ :

$$\begin{aligned} N &= \frac{n - f_w + 2p}{s_w} + 1 \\ M &= \frac{m - f_h + 2p}{s_h} + 1 \end{aligned} \tag{3.7}$$

όπου  $p$  το μέγεθος του padding γύρω από την εικόνα, και  $s_w$  και  $s_h$  το stride κατά το πλάτος και το ύψος της εικόνας αντίστοιχα. Οι τελευταίοι δύο είναι ακέραιοι αριθμοί οι οποίοι ορίζουν το βηματισμό της συνέλιξης κατά πλάτος και ύψος. Αν, για παράδειγμα,  $s_w = 1$ , τότε υπολογίζεται ένα pixel και η συνέλιξη προχωρά κατά 1 βήμα, δηλαδή πάει στο επόμενο, κατά σειρά, pixel. Αν όμως  $s_w = 2$ , η συνέλιξη θα προχωρήσει κατά 2 pixel πριν τον υπολογισμό, προσπερνώντας δηλαδή 1. Ομοίως και για το  $s_h$ , όπου όμως αυτή τη φορά ο βηματισμός θα γίνεται κατά ύψος, και όχι κατά πλάτος.

### 3.2.4 Παραδείγματα

Για να κατανοηθεί καλύτερα η αξία και η λειτουργία των φίλτρων, καθώς και η λογική με την οποία επιλέγονται οι πυρήνες, θα αναφερθούν μερικά παραδείγματα. Το αποτέλεσμα κάθε φίλτρου φαίνεται στην εικόνα 3.5, σελ. 41.

#### Box φίλτρο εξομάλυνσης (Box blur)

Η εξομάλυνση είναι μια διαδικασία η οποία χρησιμοποιείται συχνά για τη μείωση του θορύβου στις εικόνες μας. Με την εφαρμογή του, κάθε pixel της εικόνας παίρνει

τιμή ίση με το μέσο όρο όλων των γειτονικών pixels στην αρχική εικόνα. Ο πυρήνας ενός  $3 \times 3$  box blur γράφεται ως:

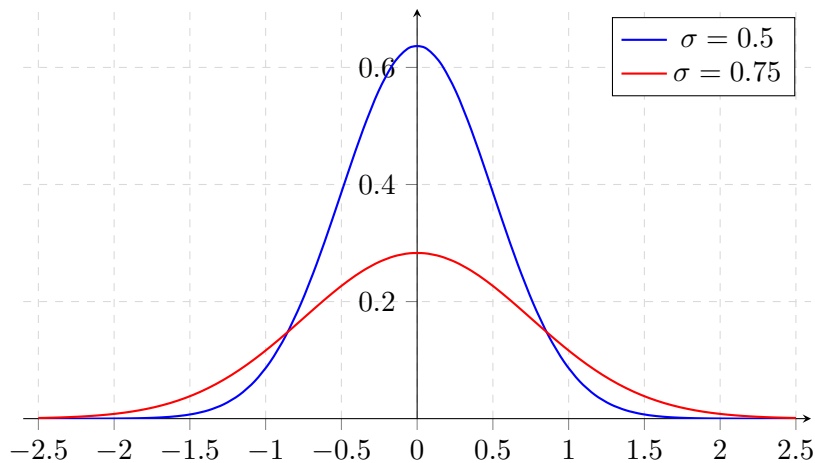
$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (3.8)$$

### Γκαουσιανό φίλτρο εξομάλυνσης (Gaussian blur)

Πρόκειται για το πιο ευρέως χρησιμοποιούμενο φίλτρο εξομάλυνσης. Μία Γκαουσιανή συνάρτηση μιας μεταβλητής με απόκλιση  $\sigma$  έχει την ακόλουθη μορφή:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (3.9)$$

Είναι η, γνωστή σε όλους μας, κανονική κατανομή με το χαρακτηριστικό σχήμα καμπάνας:



Σχήμα 3.4: Η γκαουσιανή κατανομή με  $\sigma = 0.5$  και  $\sigma = 0.75$

Αντίστοιχα, σε δύο διαστάσεις έχει τη μορφή:

$$g(x, y) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (3.10)$$

Η συνέλιξη της γκαουσιανής συνάρτησης με την εικόνα μπορεί να προσεγγιστεί, ορίζοντας έναν πυρήνα του οποίου οι τιμές προσομοιάζουν την κανονική αυτή κατανομή. Για παράδειγμα, μια προσέγγιση είναι η:

$$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (3.11)$$

Το γκαουσιανό φίλτρο εμφανίζει πολλές ενδιαφέρουσες και χρήσιμες ιδιότητες μερικές εκ των οποίων είναι:

- Τα βάρη μειώνονται ομαλά μέχρι το 0 καθώς αυξάνεται η απόσταση από το

κεντρικό pixel, καθιστώντας τις τιμές κοντά στο κέντρο πιο σημαντικές από τις τιμές οι οποίες βρίσκονται μακριά από αυτό. Η μεταβλητή  $\sigma$  είναι αυτή που ορίζει πόσο εστιασμένη ή όχι θα είναι η γειτονιά. Το 95% του συνολικού βάρους θα περιέχεται μέσα σε  $2\sigma$  από το κέντρο.

- Συμμετρία. Κατοπτρίζοντας τον πυρήνα για την εφαρμογή της συνέλιξης το φίλτρο δεν μεταβάλλεται.

### Όξυνση (sharpen)

Ας θεωρήσουμε ότι θέλουμε να αυξήσουμε την οξύτητα μιας εικόνας. Αυτό σημαίνει πως κάθε pixel στη νέα μας εικόνα θα πρέπει να ξεχωρίζει από τα γειτονικά του περισσότερο από ότι στην παλιά εικόνα. Επομένως μπορούμε να δημιουργήσουμε ένα kernel το οποίο να δίνει περισσότερο βάρος στο κεντρικό μας pixel και αρνητικό βάρος στη γειτονική περιοχή του. Ως παράδειγμα μπορούμε να δώσουμε το:

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (3.12)$$

Αν και υπάρχουν πολλοί τρόποι με τους οποίους μπορεί κάποιος να ανιχνεύσει ακμές σε μια εικόνα, εδώ θα παρουσιαστεί εν συντομία η μέθοδος Sobel. Το φίλτρο αυτό διαφέρει από τα προηγούμενα καθώς αποτελείται από 2, αντί για 1, πυρήνες. Ο ένας πυρήνας, στην πραγματικότητα, είναι απλώς ο άλλος με περιστροφή  $90^\circ$ .

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (3.13)$$

Οι πυρήνες αυτοί είναι σχεδιασμένοι έτσι ώστε να αποκρίνονται σε κάθετες και οριζόντιες ακμές. Μπορούμε να εφαρμόσουμε τον κάθε ένα ξεχωριστά και να συνδυάσουμε τα αποτελέσματα στο τέλος σύμφωνα με:

$$|G'| = |G_x| + |G_y| \quad (3.14)$$

## 3.3 Χρωματικοί χώροι

Όπως έχει εξηγηθεί και στην ενότητα 1.2.6, για να λάβουμε τη χρωματική απόχρωση σε συνδυασμό με την φωτεινότητα για ένα σημείο, μετρούμε την έξοδο των pixels για τρία διαφορετικά μήκη κύματος, το κόκκινο (R,  $\approx 700$  nm), το πράσινο (G,  $\approx 550$  nm) και το μπλε (B,  $\approx 440$  nm). Για να εξηγηθεί ο ακριβής λόγος που γίνεται αυτό, είναι απαραίτητο να κατανοηθεί ο τρόπος με τον οποίο ο άνθρωπος αντιλαμβάνεται το χρώμα.





(α') Αρχική εικόνα



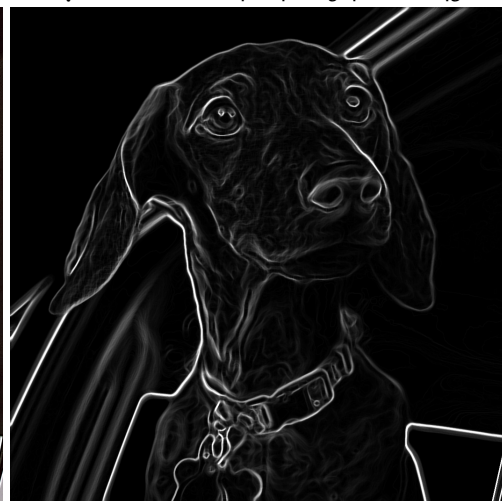
(β') Βox φίλτρο εξομάλυνσης



(γ') Γκαουσιανό φίλτρο εξομάλυνσης



(δ') Φίλτρο οξυσσης



(ε') Φίλτρο ανίχνευσης ακμών

Σχήμα 3.5: Παράδειγμα εφαρμογής φίλτρων σε μια εικόνα σκύλου.

### 3.3.1 Ανθρώπινη αντίληψη

Το φως το οποίο εκπέμπεται από πηγές ή που ανακλάται σε επιφάνειες έχει περισσότερη ή λιγότερη ενέργεια σε διαφορετικά μήκη κύματος. Αυτή η κατανομή της ενέργειας με το μήκος κύματος ονομάζεται φασματική πυκνότητα ενέργειας, και με τον όρο «χρώμα της πηγής» εννοούμε ακριβώς αυτό το φάσμα. Το ανθρώπινο οπτικό σύστημα αποκρίνεται σε φως του οποίου το μήκος κύματος κυμαίνεται περίπου μεταξύ 400nm και 700nm. Αυτό το τμήμα αποτελεί τα ορατά μήκη κύματος του φωτός και χωρίζεται σε ζώνες όπως κόκκινο, πορτοκαλί, κίτρινο, πράσινο, μπλε, μοβ.



Σχήμα 3.6: Το ορατό φάσμα φωτός.

Η αντίληψη του χρώματος στον άνθρωπο πραγματοποιείται μέσω των φωτοϋποδοχέων του ανθρώπινου ματιού, οι οποίοι βρίσκονται στον αμφιβληστροειδή χιτώνα και αποκρίνονται διαφορετικά σε κάθε μήκος κύματος του φωτός, παράγοντας ένα σήμα το οποίο ερμηνεύεται από τον άνθρωπο ως χρώμα. Υπάρχουν τέσσερις διαφορετικές φωτοχρωστικές στον αμφιβληστροειδή, μία στα ραβδία και μία σε κάθε έναν από τους τρεις τύπους κωνίων. Από αυτές, τα 3 κωνία είναι υπεύθυνα για την αίσθηση των χρωμάτων. Πιο συγκεκριμένα, η μία φωτοχρωστική εμφανίζει μέγιστο απορρόφησης στα 420nm (S-κωνία, ευαίσθητα στην ιώδη και μπλε περιοχή), η δεύτερη στα 530nm (M-κωνία, ευαίσθητα στην πράσινη και κίτρινη περιοχή) και η τρίτη στα 560nm (L-κωνία, ευαίσθητα στην κίτρινη και κόκκινη περιοχή). Το μοναδικό σύνολο αποκρίσεων για κάθε μήκος κύματος είναι αυτό που καθιστά το νευρικό σύστημα ικανό να ξεχωρίσει το μήκος κύματος που προσπίπτει στον αμφιβληστροειδή.

### 3.3.2 Περιγραφή του χρώματος

Η ακριβής περιγραφή των χρωμάτων έχει πολύ μεγάλη εμπορική σημασία. Απλά ονόματα τα οποία περιγράφουν τα χρώματα είναι ανεπαρκή, καθώς εκτός από το γεγονός ότι ελάχιστοι άνθρωποι γνωρίζουν αρκετά ονόματα χρωμάτων ώστε να περιγράψουν το χρωματικό φάσμα στο σύνολό του, μπορεί να χρησιμοποιούν ένα όνομα για να περιγράψουν μία ευρεία γκάμα χρωμάτων. Επομένως χρειάζεται ένα πρότυπο περιγραφής χρωμάτων το οποίο δεν βασίζεται σε ονόματα αλλά σε αριθμούς.

### 3.3.3 Γραμμικοί χώροι χρώματος

Αντλώντας έμπνευση από την ανθρώπινη φυσιολογία, μπορούμε να περιγράψουμε το χρώμα ως γραμμικό συνδυασμό τριών πρωταρχικών χρωμάτων (primary colors). Αν επιλέξουμε ως κύρια τρία πρωταρχικά χρώματα  $P_1(\lambda)$ ,  $P_2(\lambda)$ ,  $P_3(\lambda)$ , για παράδειγμα

το κόκκινο (R), το πράσινο (G) και το μπλέ (B), τότε ένα οποιοδήποτε χρώμα μπορεί να παρασταθεί ως γραμμικός συνδυασμός των τριών αυτών χρωμάτων:

$$S(\lambda) \stackrel{H}{=} w_1 P_1(\lambda) + w_2 P_2(\lambda) + w_3 P_3(\lambda)$$

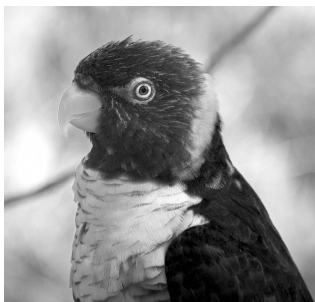
όπου οι συντελεστές  $w_1, w_2, w_3$  αποτελούν βάρη τα οποία παριστάνουν τα ποσοστά μείξης των πρωταρχικών χρωμάτων. Ο τελεστής  $\stackrel{H}{=}$  δεν αποτελεί αριθμητική ισότητα αλλά οπτικό ταίριασμα από ανθρώπινο παρατηρητή ενός τεστ χρώματος με μείγμα των πρωταρχικών.

### RGB

Ένας πολύ σημαντικός γραμμικός χώρος χρώματος είναι ο RGB. Τα τρία πρωταρχικά χρώματα, είναι, όπως δηλώνει και το όνομά του, το κόκκινο (Red - R), το πράσινο (Green - G), και το μπλε (Blue - B). Οι τρεις αυτές συνιστώσες για μια έγχρωμη εικόνα φαίνονται στο σχήμα 3.7. Δηλαδή, μία έγχρωμη RGB εικόνα αποτελείται από 3 επιμέρους μονόχρωμες εικόνες, κάθε μία αντιπροσωπεύει τις εντάσεις για κάθε πρωταρχικό χρώμα. Αυτός ακριβώς είναι και ο λόγος που, όπως αναφέρθηκε στην ενότητα 3.1.2, τα pixels μίας έγχρωμης εικόνας αποτελούνται από 24 bits αντί για τα 8 bits της μονοχρωματικής.



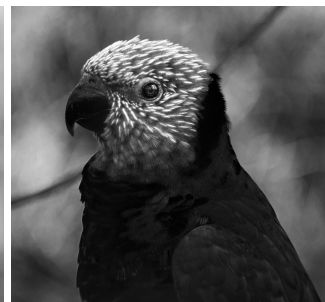
(α) Έγχρωμη εικόνα



(β) Κόκκινο (R) κανάλι



(γ) Πράσινο (G) κανάλι



(δ) Μπλε (B) κανάλι

Σχήμα 3.7: Οι RGB χρωματικές συνιστώσες.

### CIE XYZ

Ο χώρος CIE XYZ είναι επίσης ένας πολύ σημαντικός χώρος χρώματος. Για να γίνει όμως κατανοητός είναι απαραίτητη η κατανόηση του όρου CMF, (Color Matching

Functions), καθώς και ενός γεγονότος το οποίο, πολύ βολικά, αποκρύφθηκε κατά την εξήγηση του χώρου RGB.

Όταν συνδυάζουμε χρώματα με γραμμικό τρόπο, μπορούμε να καθορίσουμε πειραματικά τα βάρη, για σταθερά πρωταρχικά χρώματα, που θα πρέπει να χρησιμοποιηθούν ώστε το προκύπτον χρώμα να ταιριάζει με μια πηγή γνωστού χρώματος, ή, αλλιώς, γνωστής φασματικής πυκνότητας ενέργειας. Δηλαδή, για τα πρωταρχικά χρώματα  $P_1$ ,  $P_2$  και  $P_3$ , μπορούμε να λάβουμε, πειραματικά, *συναρτήσεις χρωματικού ταιριάσματος* (Color Matching Functions - CMFs), μεταβάλλοντας τα βάρη για κάθε μήκος κύματος έτσι ώστε να πετυχαίνουμε κάθε φορά χρωματικό ταίριασμα, και καταγράφοντας για κάθε μήκος κύματος τα αντίστοιχα σωστά βάρη. Οι πίνακες που δημιουργούνται με αυτόν τον τρόπο περιέχουν τις συναρτήσεις χρωματικού ταιριάσματος για κάθε πρωταρχικό χρώμα και συμβολίζονται ως  $f_1(\lambda)$ ,  $f_2(\lambda)$  και  $f_3(\lambda)$ . Οι 3 αυτές συναρτήσεις είναι που δίνουν τα βάρη για να ταιριάζουμε μια πηγή φωτός σε κάθε μήκος κύματος.

Δηλαδή, για ένα συγκεκριμένο μήκος κύματος  $\lambda_0$  έχουμε:

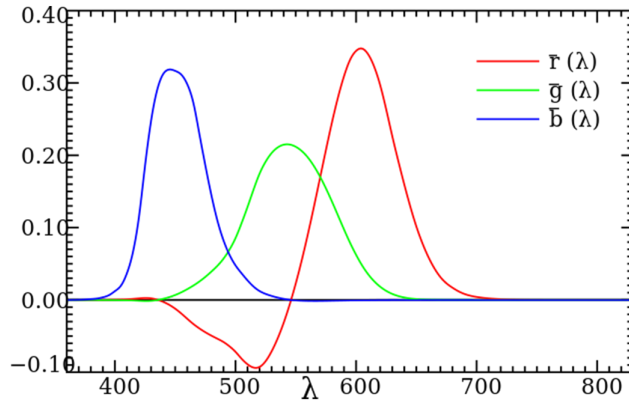
$$S(\lambda_0) \stackrel{H}{=} f_1(\lambda_0)P_1 + f_2(\lambda_0)P_2 + f_3(\lambda_0)P_3 \quad (3.15)$$

Εάν τώρα θέλουμε να ταιριάζουμε μία μη-μονοχρωματική πηγή χρώματος  $S(\lambda)$  θα πρέπει να χρησιμοποιήσουμε τα βάρη για όλα τα μήκη κύματος που περιέχει η πηγή αυτή. Επομένως:

$$\begin{aligned} S(\lambda) &\stackrel{H}{=} w_1P_1 + w_2P_2 + w_3P_3 \\ &\stackrel{H}{=} \left\{ \int_{\Lambda} f_1(\lambda)S(\lambda)d\lambda \right\} P_1 + \left\{ \int_{\Lambda} f_2(\lambda)S(\lambda)d\lambda \right\} P_2 + \left\{ \int_{\Lambda} f_3(\lambda)S(\lambda)d\lambda \right\} P_3 \end{aligned} \quad (3.16)$$

Αποδεικνύεται ότι, εάν τα πρωταρχικά χρώματα είναι αληθινά, τουλάχιστον μία από τις CMFs είναι απαραίτητο να είναι αρνητική για κάποια μήκη κύματος (σχήμα 3.8). Αυτός ο τρόπος ταιριάσματος ονομάζεται αφαιρετικός (subtractive matching). Το μη βολικό αυτό γεγονός μας οδηγεί στο να ορίσουμε συναρτήσεις χρωματικού ταιριάσματος οι οποίες είναι παντού θετικές. Τότε το ταίριασμα, σε απόλυτη αντιστοιχία με πριν, ονομάζεται προσθετικό (additive matching). Φυσικά, εφόσον έχουμε πει ότι για αληθινά χρώματα τουλάχιστον μία εκ των συναρτήσεων πρέπει να είναι κάπου αρνητική, αυτό σημαίνει ότι τα πρωταρχικά χρώματα σε αυτή την περίπτωση θα πρέπει να είναι φανταστικά.

Έτσι δημιουργήθηκε ο χώρος CIE XYZ. Ο όρος CIE προέρχεται από το όνομα της διεθνούς επιτροπής Commission internationale de l'éclairage η οποία δημιουργήθηκε για την ανάπτυξη τέτοιων προτύπων. Ο όρος XYZ προέρχεται από τις τρεις συνιστώσες  $X$ ,  $Y$ , και  $Z$ , οι οποίες έχουν επιλεγεί με τέτοιο τρόπο ώστε οι συναρτήσεις χρωματικού ταιριάσματος να είναι παντού θετικές. Οι συνιστώσες αυτές μπορούν να προκύψουν από τον ακόλουθο γραμμικό μετασχηματισμό των συνιστωσών



Σχήμα 3.8: Γραφική απεικόνιση των συναρτήσεων χρωματικού ταιριάσματος για τα πρωταρχικά χρώματα κόκκινο, πράσινο και μπλε (RGB). Η κόκκινη χρωματική συνιστώσα έχει και αρνητικές τιμές.

RGB:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3.17)$$

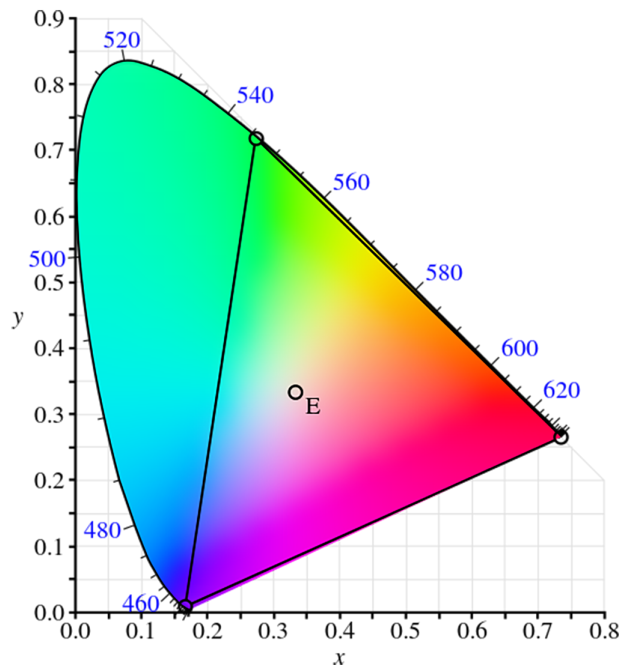
Η κανονικοποίηση των  $X$ ,  $Y$ ,  $Z$  συνιστωσών μπορεί να γίνει σύμφωνα με τη σχέση:

$$(x, y, z) = \left( \frac{X}{X+Y+Z}, \frac{Y}{X+Y+Z}, \frac{Z}{X+Y+Z} \right) \quad (3.18)$$

Οι κανονικοποιημένες αυτές συνιστώσες διαγράφουν ένα επίπεδο, αφού  $x + y + z = 1$ . Μπορούμε να δημιουργήσουμε ένα διάγραμμα αναπαριστώντας τα ανθρωπίνως αντιληπτά χρώματα στο επίπεδο των ανεξάρτητων χρωματικών συνιστωσών  $(x, y)$ . Το διάγραμμα αυτό είναι ευρέως διαδεδομένο και ονομάζεται *CIE-xy χρωματικό διάγραμμα*. Εάν σε αυτό το διάγραμμα επιλέξουμε το κόκκινο στα 700nm, το πράσινο στα 546nm και το μπλε στα 435nm, ενώνοντας τις κορυφές θα δούμε όλο το σύνολο των χρωμάτων που μπορούμε να αναπαραστήσουμε με τα 3 αυτά πρωταρχικά χρώματα. Καταλήγουμε έτσι στην επαλήθευση, και συμπέρασμα, ότι είναι αδύνατο να αναπαραστήσουμε όλα τα ανθρωπίνως αντιληπτά χρώματα χρησιμοποιώντας αποκλειστικά προσθετικό ταίριασμα. Δηλαδή, αυτό που αποφύγαμε να παρατηρήσουμε στην περιγραφή του χώρου RGB στην προηγούμενη ενότητα, είναι ότι στην πραγματικότητα δεν περιέχει όλα τα χρώματα τα οποία μπορεί να αντιληφθεί ο άνθρωπος, παρά μόνο ένα υποσύνολο (γκαμούτ) αυτών.

### 3.3.4 Μη-γραμμικοί χώροι χρώματος

Οι γραμμικοί χώροι χρώματος παρουσιάζουν αρκετά πλεονεκτήματα τα οποία προέρχονται από την αλγεβρική δομή των γραμμικών χώρων, όμως δεν αντιστοιχούν καλά στην ανθρώπινη αντίληψη χρώματος. Για παράδειγμα, οι συντεταγμένες ενός



Σχήμα 3.9: Το χρωματικό διάγραμμα και το τρίγωνο που σχηματίζουν οι RGB συνιστώσες για προσθετικό ταίριασμα.

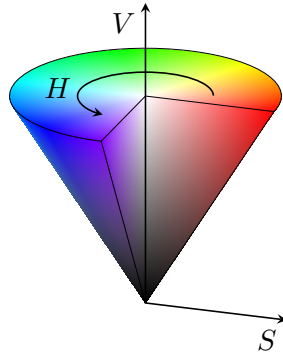
χρώματος σε ένα γραμμικό χώρο δεν περιέχουν ιδιότητες όπως είναι η *απόχρωση*, η *φωτεινότητα* και ο *κορεσμός*. Πρόκειται για σημαντικές ιδιότητες, καθώς εάν για παράδειγμα μας ενδιαφέρει να ελέγξουμε εάν ένα χρώμα ανήκει σε ένα συγκεκριμένο εύρος από κόκκινα, θα θέλαμε να μπορούμε να βρούμε απευθείας την ποσοτική μορφή της απόχρωσης του χρώματος.

Ακόμη μία δυσκολία στους γραμμικούς χώρους χρώματος είναι ότι οι συντεταγμένες ενός χρώματος δεν αντιστοιχούν στην ανθρώπινη αντίληψη της τοπολογίας των χρωμάτων. Ένας τρόπος να το αντιληφθεί κανείς αυτό είναι να φανταστεί τις τοπολογικές σχέσεις των αποχρώσεων, με την έννοια ότι το κόκκινο είναι δίπλα στο μοβ και το πορτοκαλί, το πορτοκαλί δίπλα στο κόκκινο και το κίτρινο, το κίτρινο δίπλα στο πορτοκαλί και το πράσινο, και το πράσινο δίπλα στο κίτρινο και το κυανό (σχήμα 3.6). Η διάταξη των χρωμάτων αυτών μοντελοποιείται από την κατανομή τους σε ένα κύκλο. Αυτό σημαίνει ότι δεν μπορεί να χρησιμοποιηθεί ένας γραμμικός χώρος συντεταγμένων για την περιγραφή της απόχρωσης.

## HSV

Η τυπική μέθοδος η οποία λύνει προβλήματα του είδους που αναφέρθηκαν πιο πάνω είναι η κατασκευή ενός χρωματικού χώρου ο οποίος εφαρμόζει έναν μη-γραμμικό μετασχηματισμό στο χώρο RGB. Ένας από αυτούς τους χώρους είναι και ο HSV, ο οποίος ονομάζεται έτσι από τα αρχικά των τριών συνιστωσών του, την απόχρωση (Hue), τον κορεσμό (Saturation) και τη φωτεινότητα (ή ακόμα καλύτερα *λαμπρότητα*, Value που σημαίνει Brightness). Ο χώρος αυτός προκύπτει παίρνοντας τον κυβικό χώρο συντεταγμένων του RGB, (κύβος διότι διαθέτει 3 κάθετες μεταξύ τους συνι-

σώσεις, οι οποίες κυμαίνονται μεταξύ του 0 και 1), και περιστρέφοντας τον κύβο αυτό ώστε η διαγώνιος  $(0, 0, 0) - (1, 1, 1)$  των γκρίζων χρωμάτων να γίνει ο κάθετος άξονας στον νέο τρισδιάστατο χρωματικό χώρο. Κατά μήκος αυτού του αχρωματικού άξονα ορίζουμε τη λαμπρότητα, ή αλλιώς φωτεινότητα,  $V$ . Το οριζόντιο επίπεδο κάθετο στον άξονα αποτελεί το χρωματικό επίπεδο. Η γωνία είναι η απόχρωση,  $H$ , ενώ η απόσταση από τον κάθετο άξονα είναι ο κορεσμός,  $S$ .



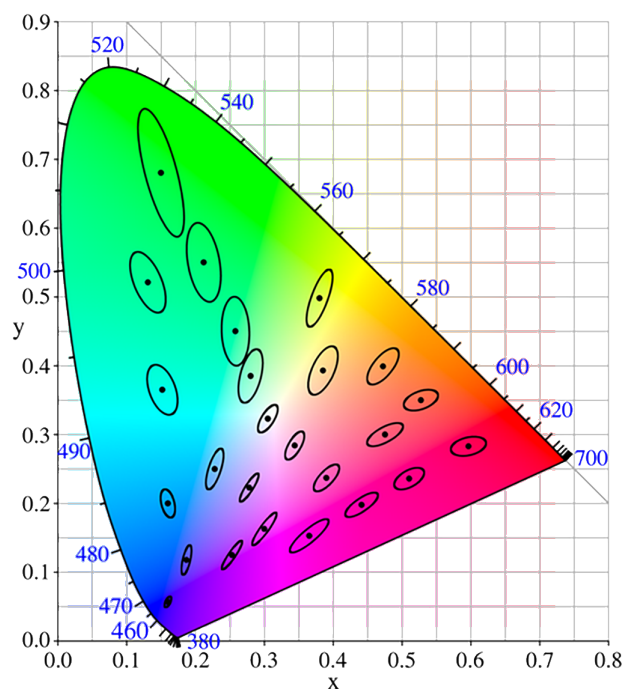
Σχήμα 3.10: Ο χρωματικός χώρος HSV ως κώνος, και οι συνιστώσες του.

### CIE LAB

Ο χώρος Lab είναι ακόμα ένας μη-γραμμικός χώρος ο οποίος όμως αναπτύχθηκε έχοντας κατά νου μια διαφορετική φιλοσοφία.

Συνήθως δεν μπορούμε να αναπαραστήσουμε τα χρώματα με απόλυτη ακρίβεια. Αυτό σημαίνει ότι είναι σημαντικό να γνωρίζουμε εάν μια χρωματική διαφορά θα γίνει αντιληπτή από τον ανθρώπινο παρατηρητή. Μπορούμε να βρούμε οριακά αντιληπτές διαφορές τροποποιώντας ένα χρώμα και δείχνοντάς το σε παρατηρητές μέχρι αυτοί να μπορέσουν να αντιληφθούν ότι το χρώμα αυτό έχει αλλάξει σε σχέση με το αρχικό χρώμα. Όταν αυτές οι διαφορές σχεδιαστούν σε ένα χρωματικό χώρο, σχηματίζουν όρια περιοχών χρώματος τα οποία είναι πανομοιότυπα, τουλάχιστον όσον αφορά ανθρώπινους παρατηρητές, με τα αρχικά χρώματα. Τα όρια αυτά αποτελούν ελλείψεις, οι οποίες ονομάζονται *ελλείψεις MacAdam* και περικλείουν το σύνολο χρωμάτων που ένας ανθρώπινος παρατηρητής θα ταίριαζε με το χρώμα των κέντρων τους [42]. Το περίγραμμά τους καθορίζει τις μόλις-διακρινόμενες διαφορές. Το γεγονός ότι παρουσιάζονται ελλείψεις και όχι κύκλοι, σημαίνει πως η διαφορά μεταξύ δύο  $(x, y)$  σημείων στον χρωματικό χώρο, η οποία δίνεται από τη σχέση  $\sqrt{(\Delta x)^2 + (\Delta y)^2}$ , είναι ένα κακό μέτρο για τη μέτρηση της διαφοράς των δύο χρωμάτων. Αντιθέτως, σε έναν *ομοιόμορφο ισοτροπικό χρωματικό χώρο* η απόσταση δύο σημείων είναι μια πιστή ένδειξη για τη διαφορά δύο χρωμάτων. Σε έναν τέτοιο χώρο, όταν η απόσταση από ένα χρώμα είναι μικρότερη από κάποιο όριο, ένας ανθρώπινος παρατηρητής αδυνατεί να καταλάβει τη χρωματική διαφορά.

Ο χώρος CIE LAB είναι πλέον ο καθολικά δημοφιλέστερος ομοιόμορφος χώρος χρώματος. Οι συνιστώσες  $(L^*, a^*, b^*)$  του χώρου αυτού προκύπτουν από το μη γραμ-



Σχήμα 3.11: Ελλείψεις MacAdam. Οι ελλείψεις αυτές είναι μεγενθυμένες κατά 10 φορές.

μικό μετασχηματισμό του χώρου XYZ:

$$\begin{aligned}
 L^* &= 116 \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 \\
 a^* &= 500 \left[ \left( \frac{X}{X_n} \right)^{\frac{1}{3}} - \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} \right] \\
 b^* &= 200 \left[ \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left( \frac{Z}{Z_n} \right)^{\frac{1}{3}} \right]
 \end{aligned}
 \tag{3.19}$$

όπου  $X_n$ ,  $Y_n$  και  $Z_n$  είναι οι  $X$ ,  $Y$  και  $Z$  συντεταγμένες ενός λευκού σημείου αναφοράς. Οι συνιστώσες  $L^*$ ,  $a^*$ ,  $b^*$ , αντιστοιχούν στη φωτεινότητα (lightness), τη διαφορά μεταξύ της κόκκινης και της πράσινης συνιστώσας, και τη διαφορά μεταξύ τη κίτρινης και της μπλε συνιστώσας αντίστοιχα. Οι τιμές της  $L^*$  κυμαίνονται μεταξύ του 0 για το μαύρο, και του 100 για το λευκό. Οι  $a^*$ ,  $b^*$  συνιστώσες παριστάνουν διαφορές μεταξύ ανταγωνιστικών χρωμάτων. Για  $a^* > 0$  έχουμε κόκκινο χρώμα, ενώ για  $a^* < 0$  πράσινο. Αντίστοιχα,  $b^* > 0$  παριστάνει κίτρινο χρώμα ενώ  $b^* < 0$  μπλε. Για  $a^* = b^* = 0$  έχουμε το λευκό αναφοράς που αναφέρθηκε παραπάνω.

Η σημαντικότητα του LAB χώρου έγκειται στο ότι είναι ιδιαίτερα ομοιόμορφος. Σε πολλά προβλήματα είναι σημαντικό να γνωρίζουμε το πόσο διαφορετικά δύο χρώματα θα φαίνονται σε έναν ανθρώπινο παρατηρητή, και οι διαφορές ανάμεσα στις LAB συντεταγμένες αποτελούν έναν καλό οδηγό.



## Κεφάλαιο 4

# Νευρωνικά Δίκτυα

Ο προγραμματισμός ηλεκτρονικών υπολογιστών περιλαμβάνει αλγορίθμους, δηλαδή πεπερασμένες ακολουθίες εντολών οι οποίες δίνονται από τον προγραμματιστή στον υπολογιστή ώστε να υλοποιηθεί κάποια συγκεκριμένη λειτουργία. Ο προγραμματιστής, δηλαδή, είναι αυτός που σκέφτεται τα βήματα για την επίλυση ενός προβλήματος, και δίνοντάς απλές, αυστηρά ορισμένες εντολές στον υπολογιστή, αυτός εκτελεί τυφλά πράξεις λύνοντας το πρόβλημα. Αυτό λειτουργεί πολύ καλά για προβλήματα στα οποία τα βήματα που απαιτούνται για την επίλυση είναι γνωστά.

Υπάρχουν όμως περιπτώσεις όπου τα βήματα που πρέπει να ακολουθηθούν δεν είναι επακριβώς γνωστά ούτε είναι εύκολο να καθοριστούν. Σε τέτοια προβλήματα είναι γνωστό το επιθυμητό αποτέλεσμα, όμως ο άνθρωπος δυσκολεύεται να δείξει στον υπολογιστή πώς να φτάσει σε αυτό. Ένα απλό παράδειγμα είναι το πρόβλημα αναγνώρισης ενός αριθμού σε μία εικόνα. Είναι αδύνατο για έναν άνθρωπο να εξηγήσει τα λογικά βήματα που ακολουθεί ώστε να αναγνωρίσει έναν αριθμό - το μόνο που γνωρίζει είναι πως έχει δει αρκετές φορές τον αριθμό, για παράδειγμα, '1', ώστε να μπορεί να τον ξεχωρίσει από τον '2'. Το ιδανικό λοιπόν θα ήταν να αρκούσε απλώς να δείξουμε στον υπολογιστή αρκετές εικόνες αριθμών μέχρι εκείνος να μάθει από μόνος του τη διαφορά μεταξύ αυτών και να μπορεί να τους ξεχωρίσει.

Άρα ένας έξυπνος υπολογιστής πρέπει να μπορεί να μαθαίνει, όπως ακριβώς και ο άνθρωπος, χωρίς να χρειάζεται να του δείξουμε το πώς, αλλά μονάχα το τί. Έτσι εισάγεται η έννοια της μηχανικής μάθησης. Ο τρόπος με τον οποίο προγραμματίζεται ένας υπολογιστής ώστε να επιτελεί αυτή τη λειτουργία αντλεί την έμπνευσή του από την ανθρώπινη φυσιολογία και ο αλγόριθμος αναφέρεται ως ένα τεχνητό νευρωνικό δίκτυο.

### 4.1 Δομή

Το βασικό δομικό στοιχείο ενός νευρωνικού δικτύου είναι ο *νευρώνας*. Ο νευρώνας μπορεί να θεωρηθεί ως απλά μία δομή η οποία περιέχει έναν αριθμό. Ο αριθμός αυτός ονομάζεται *τιμή ενεργοποίησης* και ως θεωρήσουμε ότι κυμαίνεται μεταξύ των τιμών 0 και 1. Γραφικά, ένας νευρώνας αναπαρίσταται ως κύκλος με την τιμή

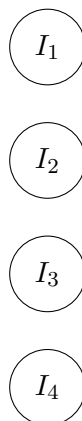
ενεργοποίησής του στο εσωτερικό του. Όσο πιο κοντά στο 1 είναι η τιμή, τόσο περισσότερο *ενεργοποιημένος* είναι ο νευρώνας και αντίστροφα.



Σχήμα 4.1: Ένας νευρώνας με τιμή ενεργοποίησης 0.7

Ένα από τα πιο γνωστά προβλήματα κατηγοριοποίησης είναι αυτό της ίριδας. Η ίριδα είναι ένα ανθοφόρο φυτό, αρκετά είδη του οποίου υπάρχουν και στη χώρα μας. Η κατηγοριοποίηση του είδους κάθε τέτοιου φυτού μπορεί να γίνει, εκτός από την οπτική παρατήρηση, και μέσω διαφόρων μετρήσεων, πιο συγκεκριμένα του πλάτους και του μήκους των πετάλων και σεπάλων του λουλουδιού.

Στη συνέχεια, οι 4 αυτές τιμές θα πρέπει να δοθούν ως είσοδος στο νευρωνικό δίκτυο το οποίο και θα κατηγοριοποιήσει το λουλούδι. Κάθε είσοδος μπορεί να αναπαρασταθεί ως μια ομάδα νευρώνων με τιμές ενεργοποίησης τις  $I_1, I_2, I_3, I_4$ . Η πρώτη αυτή ομάδα νευρώνων αποτελεί και το πρώτο επίπεδο (layer) του νευρωνικού δικτύου.

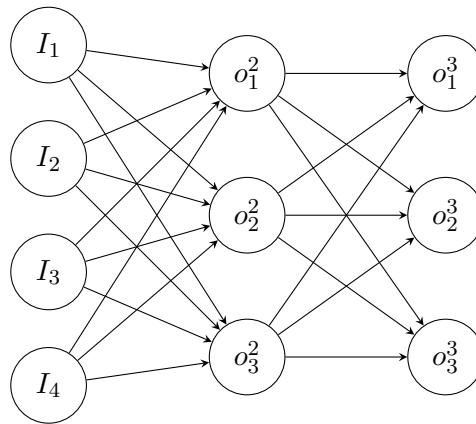


Σχήμα 4.2: Οι εισοδοί του νευρωνικού δικτύου με τη μορφή νευρώνων

Αυτοί οι νευρώνες μπορούν να συνδεθούν και με άλλους, δηλαδή με ένα άλλο επίπεδο νευρώνων, και εδώ είναι που τα νευρωνικά δίκτυα αποκτούν ενδιαφέρον: ο κάθε νευρώνας μπορεί και αλληλεπιδρά με τους υπόλοιπους στους οποίους είναι συνδεδεμένος έτσι ώστε άλλους να τους ενεργοποιεί λιγότερο, και άλλους περισσότερο. Αυτοί με τη σειρά τους ενεργοποιούν, ο καθένας ξεχωριστά, τους νευρώνες του επόμενου επιπέδου, κ.ο.κ. Τα επιπλέον επίπεδα τα οποία εισάγονται ονομάζονται «κρυφά» και το πλήθος τους, καθώς και το πλήθος των νευρώνων από τους οποίους αποτελούνται, επιλέγεται εμπειρικά καθώς διαφέρει για κάθε πρόβλημα. Στο εξής, η τιμή ενεργοποίησης ενός νευρώνα  $i$  στο επίπεδο  $k$  θα συμβολίζεται ως:  $o_i^k$ .

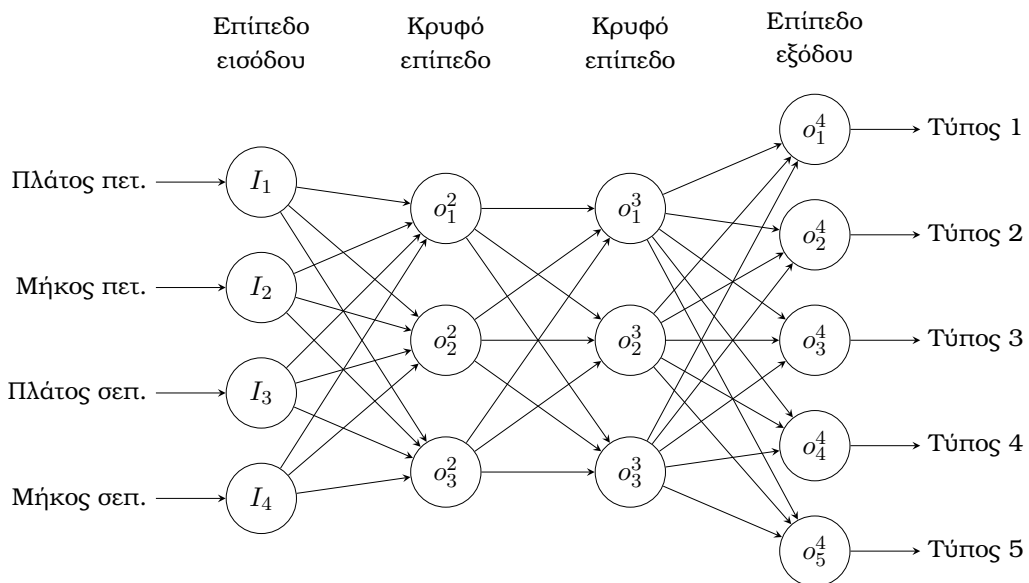
Όταν τελειώσουν τα κρυφά επίπεδα, προστίθεται ακόμα ένα, τελευταίο, επίπεδο το οποίο ονομάζεται επίπεδο εξόδου. Αν στο παράδειγμά μας θέλουμε να κατατάξουμε το λουλούδι μας σε 1 μεταξύ 5 διαφορετικών τύπων λουλουδιών, τότε η έξοδος του νευρωνικού δικτύου θα πρέπει να περιλαμβάνει 5 νευρώνες, έναν για

Επίπεδο εισόδου      Κρυφό επίπεδο      Κρυφό επίπεδο



Σχήμα 4.3: Προσθήκη κρυφών επιπέδων

κάθε διαφορετικό τύπο λουλουδιού. Οι νευρώνες αυτοί, εν τέλει, θα έχουν όλοι από μία τιμή ενεργοποίησης. Αυτός, όμως, ο οποίος θα είναι περισσότερο ενεργοποιημένος, δηλαδή η τιμή ενεργοποίησής του θα είναι πιο κοντά στο 1, θα είναι και αυτός στον οποίο αντιστοιχεί ο τύπος του λουλουδιού του οποίου οι μετρήσεις δόθηκαν ως είσοδοι.



Σχήμα 4.4: Το πλήρες νευρωνικό δίκτυο.

Δίνοντας λοιπόν 4 εισόδους, οι νευρώνες αλληλεπιδράσαν μεταξύ τους με τέτοιο τρόπο που τελικά μας έδωσαν ως έξοδο τον τύπο του λουλουδιού. Με παρόμοιο τρόπο είναι δυνατό ένα νευρωνικό δίκτυο να επιτελεί πιο σύνθετες λειτουργίες, όπως είναι, για παράδειγμα, η αναγνώριση και κατηγοριοποίηση μιας εικόνας ή ενός αρχείου βίντεο.

Για να κατανοηθεί όμως πλήρως η λειτουργία του νευρωνικού δικτύου, θα πρέπει

να εξηγηθούν ακριβώς τα μέρη από τα οποία αποτελείται ο κάθε νευρώνας και πώς αλληλεπιδρά με άλλους με τους οποίους είναι συνδεδεμένος.

#### 4.1.1 Το μοντέλο ενός νευρώνα

Στις προηγούμενες παραγράφους δεν δόθηκε ένας αυστηρός ορισμός για τον τεχνητό νευρώνα - θεωρήθηκε απλά ότι πρόκειται για μία δομή η οποία περιέχει έναν αριθμό. Ο αριθμός αυτός στην πραγματικότητα αποτελεί απλά την έξοδο του νευρώνα, ενώ ο ίδιος ο νευρώνας είναι μία μαθηματική συνάρτηση η οποία αποτελεί ένα απλουστευμένο μοντέλο ενός βιολογικού νευρώνα. Η βασική λειτουργία του τεχνητού νευρώνα συνίσταται στη λήψη πολλαπλών σημάτων ως εισόδους, τον πολλαπλασιασμό κάθε εισόδου με ένα βάρος, την άθροιση των σημάτων αυτών και την είσοδο του αποτελέσματος σε μία μη-γραμμική συνάρτηση η οποία δίνει και το τελικό αποτέλεσμα.

Ένας νευρώνας αποτελείται, λοιπόν, από 3 βασικά μέρη:

- Τις συνάψεις, μέσω των οποίων συνδέεται με άλλους νευρώνες, και οι οποίες χαρακτηρίζονται από μία τιμή βάρους, η οποία είναι ξεχωριστή για την κάθε μία.
- Έναν αθροιστή ο οποίος αθροίζει τις εξόδους όλων των συνάψεων, τα σήματα των οποίων δέχεται ως εισόδους ο νευρώνας.
- Τη μη-γραμμική συνάρτηση ενεργοποίησης η οποία ορίζει την τελική τιμή ενεργοποίησης του νευρώνα και την φράσσει μεταξύ δύο πεπερασμένων τιμών (στο παραπάνω παράδειγμα 0 και 1).

#### Οι συνάψεις

Στους νευρώνες βιολογικών οργανισμών οι συνάψεις αποτελούν δομικές και λειτουργικές μονάδες οι οποίες συνδέουν τους νευρώνες μεταξύ τους και μεταφέρουν σήματα μεταξύ αυτών. Οι περισσότεροι νευρώνες κωδικοποιούν τις εξόδους τους ως μία απότομη μεταβολή στο ηλεκτρικό δυναμικό τους, γνωστή ως δυναμικό δράσης. Το δυναμικό δράσης προκαλείται στο σώμα του κάθε νευρώνα, και διαδίδεται στο νευρικό σύστημα μέσω των συνάψεων.

Αναλογικά με αυτόν τον τρόπο λειτουργούν και οι συνάψεις σε ένα τεχνητό νευρικό δίκτυο καθώς αποτελούν τους συνδέσμους μεταξύ των διαφορετικών νευρώνων. Πιο συγκεκριμένα, αν θεωρήσουμε μία τυχαία σύναψη η οποία συνδέει τον νευρώνα  $j$  στο επίπεδο  $k - 1$  με τον νευρώνα  $i$  στο επίπεδο  $k$ , και το σήμα εξόδου του νευρώνα  $j$  στο επίπεδο  $k - 1$  ως  $o_j^{k-1}$ , τότε το σήμα αυτό θα πολλαπλασιαστεί με το βάρος της σύναψης  $w_{ji}^k$ , και η είσοδος που θα λάβει ο νευρώνας  $i$  στο επίπεδο  $k$  θα είναι  $w_{ji}^k o_j^{k-1}$ .

## Ο αθροιστής

Ένας νευρώνας μπορεί να λάβει ταυτόχρονα πολλά διαφορετικά σήματα, το καθένα με διαφορετικό βάρος, τα οποία πρέπει να χρησιμοποιήσει για να παράγει την έξοδό του. Τα σήματα αυτά, επομένως, αθροίζονται μεταξύ τους, με την είσοδό τους στον νευρώνα, μέσω του αθροιστή.

## Η συνάρτηση ενεργοποίησης

Η συνάρτηση ενεργοποίησης ορίζει την έξοδο του νευρώνα και φράσσει την έξοδο του αθροιστή μέσα σε ένα καθορισμένο όριο τιμών. Είναι απαραίτητα μη-γραμμική. Ως παραδείγματα μπορούμε να δώσουμε 2 τέτοιες συναρτήσεις.

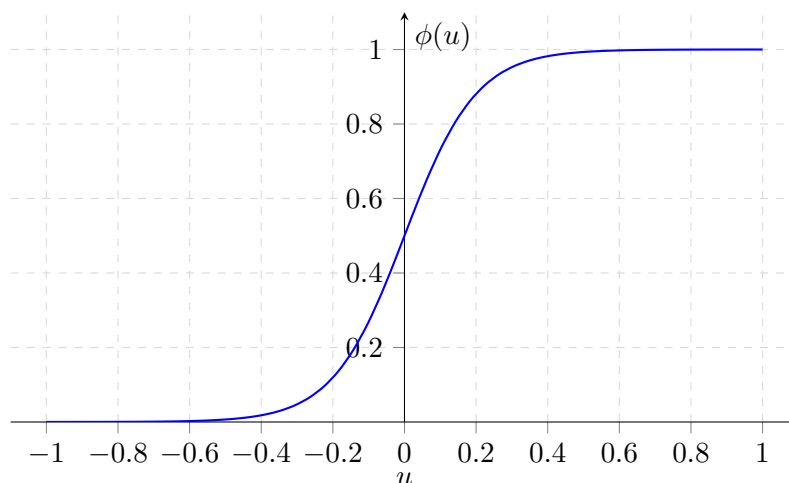
- Συνάρτηση κατωφλίου. Η συνάρτηση αυτή είναι η γνωστή συνάρτηση Heaviside και ορίζεται ως:

$$\phi(u) = \begin{cases} 1, & \text{εάν } u \geq 0 \\ 0, & \text{εάν } u < 0 \end{cases} \quad (4.1)$$

- Σιγμοειδής συνάρτηση. Πρόκειται για την πιο κοινή συνάρτηση ενεργοποίησης που χρησιμοποιείται στα νευρωνικά δίκτυα. Ορίζεται ως μία αύξουσα συνάρτηση η οποία πετυχαίνει μία ισορροπία μεταξύ της γραμμικής και της μη-γραμμικής συμπεριφοράς. Ένα παράδειγμα τέτοιας συνάρτησης είναι η:

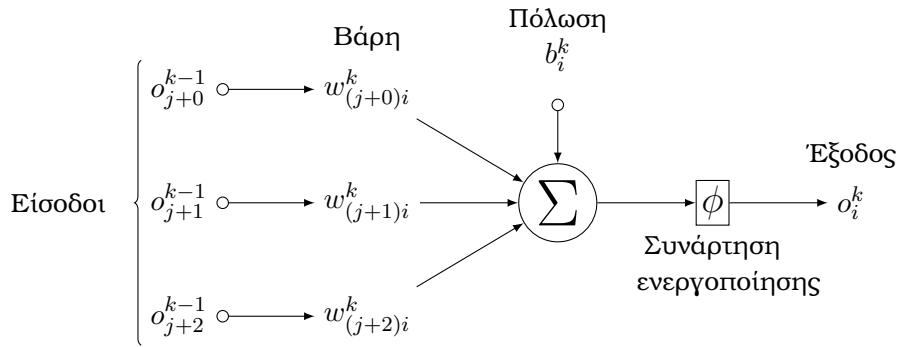
$$\phi(u) = \frac{1}{1 + \exp(-au)} \quad (4.2)$$

όπου  $a$  είναι η παράμετρος κλίσης της συνάρτησης.



Σχήμα 4.5: Η σιγμοειδής συνάρτηση με  $\alpha = 10$

Είναι σημαντικό να αναφερθεί ότι πολλές φορές χρησιμοποιείται και μία τιμή πόλωσης,  $b_i^k$  η οποία προστίθεται στην έξοδο του αθροιστή και το αποτέλεσμα είναι αυτό που δίνεται ως τιμή στην συνάρτηση ενεργοποίησης.



Σχήμα 4.6: Γραφική περιγραφή ενός νευρώνα

Μαθηματικά, η έξοδος ενός νευρώνα  $i$  στο επίπεδο  $k$  περιγράφεται, δηλαδή, από την εξίσωση:

$$o_i^k = \phi \left( \sum_{j=1}^{r_{k-1}} w_{ji}^k o_j^{k-1} + b_i^k \right) \quad (4.3)$$

όπου  $o_j^{k-1}$  είναι το σήμα εξόδου του νευρώνα  $j$  στο επίπεδο  $k - 1$  και το οποίο συνδέεται ως είσοδος στις συνάψεις,  $w_{ji}^k$  το βάρος της σύνδεσης που συνδέει τον νευρώνα  $j$  του επιπέδου  $k - 1$  με τον νευρώνα  $i$  του επιπέδου  $k$ ,  $b_i^k$  η τιμή της πόλωσης του νευρώνα  $i$  στο επίπεδο  $k$ ,  $\phi$  η συνάρτηση ενεργοποίησης και  $r_{k-1}$  το πλήθος των νευρώνων του επιπέδου  $k - 1$ .

## 4.2 Εκπαίδευση νευρωνικών δικτύων

Μία σημαντική ερώτηση που μπορεί να έχει ήδη σκεφθεί ο αναγνώστης, είναι το πώς θα ορίσουμε τα βάρη των συνάψεων καθώς και την πόλωση σε κάθε νευρώνα ώστε το τεχνητό νευρωνικό δίκτυο να επιτελεί την επιθυμητή λειτουργία.

### 4.2.1 Κανόνες εκμάθησης

Θα αρχίσουμε ορίζοντας γενικά τους κανόνες εκμάθησης. Με τον όρο αυτό, εννοούμε την διαδικασία σύμφωνα με την οποία μεταβάλλουμε τα βάρη και τις πόλωσης ενός νευρωνικού δικτύου. Ένας εναλλακτικός όρος για τη διαδικασία αυτή είναι ο «αλγόριθμος εκπαίδευσης». Ο σκοπός του κανόνα εκμάθησης είναι να εκπαιδεύσει το νευρωνικό δίκτυο έτσι ώστε αυτό να μπορεί να επιτελέσει κάποια εργασία. Οι κανόνες μπορούν να χωριστούν σε 3 κατηγορίες:

- Επιβλεπόμενη μάθηση (supervised learning)
- Μη-επιβλεπόμενη μάθηση (unsupervised learning)
- Ενισχυτική μάθηση (reinforcement learning)

## Επιβλεπόμενη μάθηση

Στην επιβλεπόμενη μάθηση, ο κανόνας εκμάθησης δίνεται με ένα σύνολο από παραδείγματα, τα δεδομένα εκπαίδευσης (training set ή dataset), τα οποία ορίζουν τη σωστή λειτουργία του νευρωνικού δικτύου:

$$\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_Q, t_Q\} \quad (4.4)$$

όπου  $p_q$  είναι η είσοδος στο δίκτυο και  $t_q$  είναι η αντίστοιχη (επιθυμητή) έξοδος. Καθώς εφαρμόζονται εισόδους  $p$  στο δίκτυο, οι αντίστοιχοι έξοδοι του δικτύου  $\hat{t}$  συγκρίνονται με τις επιθυμητές εξόδους  $t$ . Ο κανόνας εκμάθησης χρησιμοποιείται για να μεταβάλλει τα βάρη και τις πολώσεις του δικτύου έτσι ώστε να μετακινήσει τις εξόδους πιο κοντά στις επιθυμητές.

## Ενισχυτική μάθηση

Η ενισχυτική μάθηση μοιάζει αρκετά με την επιβλεπόμενη μάθηση, διαφοροποιείται όμως ως προς το ότι αντί να δίνεται κάθε φορά η σωστή έξοδος για την αντίστοιχη είσοδο, δίνεται ένας βαθμός. Ο βαθμός (grade ή score) είναι ένα μέτρο της απόδοσης του δικτύου για μία αλληλουχία εισόδων. Πρόκειται για έναν τύπο εκμάθησης ο οποίος χρησιμοποιείται πολύ πιο σπάνια από την επιβλεπόμενη μάθηση, και βρίσκει εφαρμογή κυρίως σε προβλήματα συστημάτων ελέγχου.

## Μη-επιβλεπόμενη μάθηση

Στη μη-επιβλεπόμενη μάθηση, τα βάρη και οι πολώσεις μεταβάλλονται αποκλειστικά βάσει των εισόδων στο δίκτυο. Δεν υπάρχουν επιθυμητές εξόδους. Αυτό μπορεί να φανεί παράδοξο, καθώς η λογική υποδεικνύει πως είναι αδύνατο να εκπαιδευτεί ένα δίκτυο αν δεν είναι γνωστό το τελικό αποτέλεσμα που θέλουμε να επιτύχουμε. Οι περισσότεροι όμως από αυτούς του αλγορίθμους επιτελούν απλά μία ομαδοποίηση (clustering) των δεδομένων. Μαθαίνουν, δηλαδή, να κατηγοριοποιούν τις εισόδους σε ένα πεπερασμένο αριθμό από ομάδες, ή αλλιώς, κλάσεις.

Στη συνέχεια θα επεξηγηθούν τα βασικά της επιβλεπόμενης μάθησης, τόσο λόγω της ευρείας χρήσης της, όσο και λόγω του ότι χρησιμοποιείται στα συνελκτικά νευρωνικά δίκτυα τα οποία και χρησιμοποιήθηκαν σε αυτή τη διπλωματική εργασία.

### 4.2.2 Αλγόριθμος απότομης καθόδου

Ας επανέλθουμε στο παράδειγμα της κατηγοριοποίησης της ίριδας και ας θεωρήσουμε πως χρησιμοποιούμε την αρχιτεκτονική του δικτύου της εικόνας 4.4.

Συνήθως, όταν δημιουργούμε ένα νευρωνικό δίκτυο από το μηδέν, τα βάρη και οι πολώσεις για κάθε νευρώνα επιλέγονται τυχαία μέσα από ένα σύνολο τιμών. Όπως είναι φυσικό, η έξοδος ενός τέτοιου δικτύου είναι λανθασμένη και απολύτως άχρηστη,

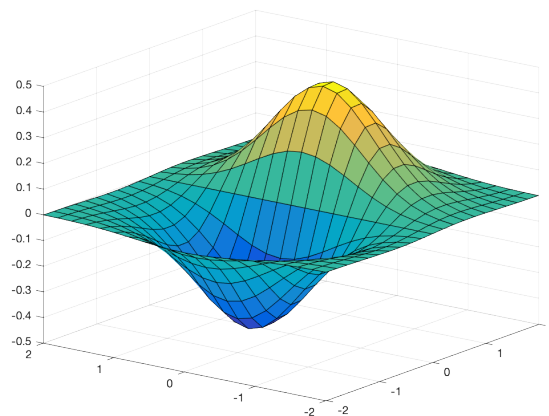
αφού πρόκειται για ένα αποτέλεσμα που προέκυψε μέσα από τυχαίες τιμές.

Θεωρούμε ένα σύνολο δεδομένων το οποίο αποτελείται από  $n$  λουλούδια, καθένα εκ των οποίων ανήκει σε 1 μεταξύ  $m$  τύπων λουλουδιών, και ο οποίος είναι γνωστός εκ των προτέρων. Το *training set* μας λοιπόν αποτελείται από  $n$  ζεύγη 4-διάστατων διανυσμάτων  $\{(\mathbf{p}_1, \mathbf{t}_1), \dots, (\mathbf{p}_n, \mathbf{t}_n)\}$ . Ορίζουμε μία *συνάρτηση κόστους* (cost function) η οποία ορίζει το πόσο λάθος είναι η έξοδος που δίνει το δίκτυό μας:

$$E = \frac{1}{2} \sum_{d=1}^n \|\hat{\mathbf{t}}_d - \mathbf{t}_d\|^2 \quad (4.5)$$

όπου  $\hat{\mathbf{t}}_d$  είναι η έξοδος που προκύπτει από το νευρωνικό δίκτυο αφού δοθεί ως είσοδος ένα  $\mathbf{p}_d$ .

Επομένως, η τιμή του σφάλματος  $E$  είναι μικρή όταν το δίκτυο δίνει αποτελέσματα κοντά στα επιθυμητά, ενώ αντιθέτως, είναι μεγάλη όταν τα επιθυμητά αποτελέσματα έχουν μεγάλη απόκλιση από αυτά που προέκυψαν. Το πρόβλημα στην εκπαίδευση του νευρωνικού δικτύου είναι η εύρεση βαρών και πολώσεων τα οποία *ελαχιστοποιούν* τη συνάρτηση κόστους. Πρόκειται, δηλαδή, για ένα πρόβλημα εύρεσης ελαχίστου. Ιδανικά θα θέλαμε να βρούμε το ολικό ελάχιστο, κάτι όμως το οποίο αποδεικνύεται εξαιρετικά δύσκολο. Συνήθως αρκεί να βρούμε ένα τοπικό ελάχιστο της συνάρτησης.



Σχήμα 4.7: Γραφική παράσταση τρισδιάστατης συνάρτησης

Μπορούμε να φανταστούμε μια τρισδιάστατη συνάρτηση ως μια έκταση γης η οποία περιέχει κοιλάδες, λόφους και πεδιάδες. Αν κάποιος βρισκόταν σε ένα τυχαίο σημείο της γης αυτής και ήθελε να κατεβεί σε όσο χαμηλότερο υψόμετρο μπορούσε, μια προφανής λύση θα ήταν η εύρεση της πλησιέστερης κατηφόρας. Η κατηφόρα αυτή σίγουρα θα οδηγεί σε κάποιο τοπικό ελάχιστο.

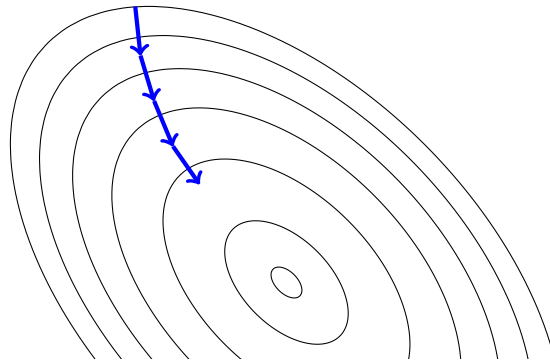
Μαθηματικά, η κλίση (gradient) μιας συνάρτησης ( $\nabla C$ ) μας δίνει την κατεύθυνση της τάχιστης ανόδου, δηλαδή, για το παράδειγμά μας, την κατεύθυνση της ανηφόρας. Αντίστοιχα, λαμβάνοντας την αρνητική τιμή της κλίσης βρίσκουμε την



κατεύθυνση της τάχιστης καθόδου, ή αλλιώς, την διεύθυνση της κατηφόρας.

$$\nabla C = \frac{\partial C}{\partial x} \mathbf{i} + \frac{\partial C}{\partial y} \mathbf{j} \quad (4.6)$$

Ο αλγόριθμος, λοιπόν, για την εύρεση ενός τοπικού ελαχίστου συνίσταται στον υπολογισμό του gradient της συνάρτησης για ένα σημείο  $\nabla C(x, y)$ , στην μετακίνηση προς την κατεύθυνση του  $-\nabla C(x, y)$  κατά ένα μικρό βήμα, και ο εκ νέου υπολογισμός του gradient. Τα βήματα αυτά επαναλαμβάνονται μέχρις ότου δεν υπάρχει πλέον άλλο σημείο κατάβασης. Η λογική αυτή μπορεί να γενικευθεί για οποιαδήποτε συνάρτηση, οποιωνδήποτε διαστάσεων.



Σχήμα 4.8: Γραφική απεικόνιση της απότομης καθόδου

Αν φανταστούμε πως κάθε βάρος του δικτύου αντιστοιχεί και σε μία τιμή, ή συνιστώσα, του διανύσματος του gradient για τη συνάρτηση κόστους, το πρόσημο της τιμής αυτής μας δείχνει σε ποιά κατεύθυνση πρέπει να αλλαχθεί ο αριθμός ώστε να βρεθεί ένα τοπικό ελάχιστο, ενώ το μέτρο της τιμής μας δείχνει πόσο θα επηρεαστεί το τελικό αποτέλεσμα από την μεταβολή αυτού του βάρους ή πόλωσης. Μπορούμε, δηλαδή, να ελαχιστοποιήσουμε τη συνάρτηση κόστους  $E$ , σύμφωνα με μια επαναληπτική διαδικασία απότομης καθόδου χρησιμοποιώντας το gradient:

$$\nabla E = \left( \frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_l} \right) \quad (4.7)$$

και ανανεώνοντας κάθε βάρος με τη σχέση:

$$\Delta w_i = -\gamma \frac{\partial E}{\partial w_i} \quad \text{για } i = 1, \dots, l \quad (4.8)$$

όπου  $\gamma$  είναι μία σταθερά εκμάθησης, δηλαδή μία παράμετρος αναλογίας που ορίζει το μήκος του βήματος προς την αρνητική κατεύθυνση του gradient σε κάθε επανάληψη.

Φυσικά, καθώς η συνάρτηση κόστους είναι κατασκευασμένη έτσι ώστε να μας δείχνει την απόκλιση των εξόδων του δικτύου από τις επιθυμητές εξόδους, το δίκτυο, καθώς οι τιμές αλλάζουν όπως υπαγορεύει το gradient, θα αρχίσει να αποκτά την επιθυμητή συμπεριφορά και να συγκλίνει προς τις επιθυμητές τιμές καθώς η συνάρτηση κόστους μειώνεται. Ο αλγόριθμος με τον οποίο υπολογίζεται αποδοτικά το

gradient ονομάζεται *backpropagation* (οπισθοδιάδοση) και βρίσκεται στην καρδιά της εκπαίδευσης των νευρωνικών δικτύων.

### 4.2.3 Backpropagation

Η μέθοδος της οπισθοδρόμησης, ευρύτερα γνωστή ως *backpropagation*, ονομάζεται έτσι διότι ο υπολογισμός του gradient πραγματοποιείται διατρέχοντας ανάποδα το νευρωνικό δίκτυο, με το gradient των βαρών του τελευταίου επιπέδου να υπολογίζεται πρώτο, και το gradient των βαρών του πρώτου επιπέδου να υπολογίζεται τελευταίο. Αυτή η ανάποδη ροή πληροφορίας επιτρέπει τον αποδοτικό υπολογισμό του gradient σε κάθε επίπεδο εν αντιθέσει με την, πιθανότατα πιο προφανή, λύση, όπου το gradient υπολογίζεται ξεχωριστά για κάθε επίπεδο.

Η λειτουργία της μεθόδου του *backpropagation* στηρίζεται κυρίως στον πολύ γνωστό κανόνα της αλυσίδας. Σύμφωνα με αυτόν, θεωρώντας τη συνάρτηση  $z = f(y(x))$  ισχύει ότι:

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx} \quad (4.9)$$

Ο κανόνας αυτός μπορεί να γενικευθεί και για περισσότερες συναρτήσεις.

Άμεση συνέπεια αυτού είναι ότι ασυνεχείς συναρτήσεις, όπως η συνάρτηση *heaviside* που έχει ήδη αναφερθεί πιο πάνω, δεν μπορούν να χρησιμοποιηθούν καθώς είναι μη-διαφορίσιμες.

Προκειμένου να απλοποιηθούν οι σχέσεις που θα παρουσιαστούν στη συνέχεια, στο εξής θα θεωρούμε πως οι πολώσεις  $b_i^k$  για έναν νευρώνα  $i$  σε ένα επίπεδο (layer)  $k$  ενσωματώνονται στα βάρη ως  $w_{0i}^k$ . Επομένως θα ισχύει η ισοδυναμία:

$$o_i^k = \phi\left(b_i^k + \sum_{j=1}^{r_{k-1}} w_{ji}^k o_j^{k-1}\right) = \phi\left(\sum_{j=0}^{r_{k-1}} w_{ji}^k o_j^{k-1}\right) \quad (4.10)$$

Όπως έχει ήδη αναφερθεί πιο πάνω, η μέθοδος *backpropagation* επιχειρεί να ελαχιστοποιήσει μια συνάρτηση κόστους:

$$E = \frac{1}{2} \sum_{d=1}^n \|\hat{t}_d - t_d\|^2 \quad (4.11)$$

Αυτό επιτυγχάνεται υπολογίζοντας, για κάθε βάρος  $w_{ji}^k$ , την τιμή  $\frac{\partial E}{\partial w_{ji}^k}$ . Καθώς η συνάρτηση κόστους μπορεί να αναλυθεί στο άθροισμα μεμονωμένων τιμών σφάλματος για κάθε ζευγάρι εισόδου-εξόδου  $\{p, t\}$ , η παράγωγος μπορεί να υπολογιστεί ξεχωριστά για κάθε ζευγάρι και ύστερα να γίνει η άθροισή τους.

Ο αλγόριθμος του *backpropagation* ξεκινά εφαρμόζοντας τον κανόνα της αλυσίδας στη μερική παράγωγο της συνάρτησης κόστους:

$$\frac{\partial E}{\partial w_{ji}^k} = \frac{\partial E}{\partial \alpha_i^k} \frac{\partial \alpha_i^k}{\partial w_{ji}^k} \quad (4.12)$$

όπου  $\alpha_i^k$  είναι η ενεργοποίηση του νευρώνα  $i$  στο επίπεδο  $k$  πριν αυτή περάσει από τη μη-γραμμική συνάρτηση ενεργοποίησης  $\phi$ . Δηλαδή,

$$\alpha_i^k = \sum_{j=0}^{r_{k-1}} w_{ji}^k o_j^{k-1} \quad (4.13)$$

Το συμπέρασμα που μπορούμε να εξαγάγουμε από την εξίσωση 4.12 είναι ότι η μεταβολή στη συνάρτηση κόστους λόγω της μεταβολής ενός βάρους ισούται με το γινόμενο της μεταβολής στο κόστος  $E$  λόγω της ενεργοποίησης  $\alpha_i^k$ , με τη μεταβολή στην ενεργοποίηση  $\alpha_i^k$  λόγω του βάρους  $w_{ji}^k$ .

Ο πρώτος όρος συνήθως αποκαλείται *σφάλμα* και συμβολίζεται ως:

$$\delta_i^k \equiv \frac{\partial E}{\partial \alpha_i^k} \quad (4.14)$$

Ο δεύτερος όρος υπολογίζεται από τη σχέση για το  $\alpha_i^k$  που έχει ήδη δοθεί:

$$\frac{\partial \alpha_i^k}{\partial w_{ji}^k} = \frac{\partial}{\partial w_{ji}^k} \sum_{j=0}^{r_{k-1}} w_{ji}^k o_j^{k-1} = o_j^{k-1} \quad (4.15)$$

Καταλήγουμε, τελικά, στο συμπέρασμα:

$$\frac{\partial E}{\partial w_{ji}^k} = \delta_i^k o_j^{k-1} \quad (4.16)$$

Δηλαδή, η μερική παράγωγος της συνάρτησης κόστους ως προς ένα βάρος ισούται με το γινόμενο του σφάλματος  $\delta_i^k$  για τον νευρώνα  $i$  στο επίπεδο  $k$ , με την έξοδο  $o_j^{k-1}$  του νευρώνα  $j$  στο επίπεδο  $k-1$ . Αυτό έχει λογική αν αναλογιστεί κανείς πως το βάρος  $w_{ji}^k$  συνδέει την έξοδο του νευρώνα  $j$  στο επίπεδο  $k-1$  με την είσοδο του νευρώνα  $i$  στο επίπεδο  $k$ .

Ξεκινώντας, λοιπόν, από το τελικό επίπεδο, και θεωρώντας πως έχουμε το δίκτυο της εικόνας 4.4, η μέθοδος backpropagation επιχειρεί να ορίσει το σφάλμα  $\delta_i^4$ , όπου 4 είναι ο αριθμός του τελικού επιπέδου και  $i = 1, \dots, 5$  ο αριθμός ενός από τους νευρώνες της εξόδου. Η μερική παράγωγος της συνάρτησης κόστους για ένα βάρος  $w_{ji}^4$  θα είναι απλά:

$$\frac{\partial E}{\partial w_{ji}^4} = \delta_i^4 o_j^3 \quad (4.17)$$

Προχωρώντας στο επόμενο επίπεδο, το επίπεδο 3, ο κανόνας της αλυσίδας χρησιμοποιείται και πάλι για τον υπολογισμό των μερικών παραγώγων. Για κάθε επίπεδο από εδώ και στο εξής, δηλαδή για  $1 \leq k < m$  ισχύει ότι:

$$\delta_i^k = \frac{\partial E}{\partial \alpha_i^k} = \sum_{l=1}^{r_{k+1}} \frac{\partial E}{\partial \alpha_l^{k+1}} \frac{\partial \alpha_l^{k+1}}{\partial \alpha_i^k} = \sum_{l=1}^{r_{k+1}} \delta_l^{k+1} \frac{\partial \alpha_l^{k+1}}{\partial \alpha_i^k} \quad (4.18)$$

και χρησιμοποιώντας την εξίσωση 4.13

$$\alpha_l^{k+1} = \sum_{j=1}^{r_k} w_{jl}^{k+1} \phi(\alpha_j^k) \quad (4.19)$$

οδηγούμαστε στη σχέση :

$$\frac{\partial \alpha_l^{k+1}}{\partial \alpha_i^k} = w_{li}^{k+1} \phi'(\alpha_i^k) \quad (4.20)$$

Συνδυάζοντας τις 4.18 και 4.20 προκύπτει η *backpropagation* εξίσωση:

$$\delta_i^k = \phi'(\alpha_i^k) \sum_{l=1}^{r_{k+1}} w_{il}^{k+1} \delta_l^{k+1} \quad (4.21)$$

Αυτή η τελευταία εξίσωση μας αποδεικνύει ότι το σφάλμα  $\delta_i^k$  στο επίπεδο  $k$  εξαρτάται από τα σφάλματα  $\delta_l^{k+1}$  του επιπέδου  $k + 1$ . Επομένως, τα σφάλματα ρέουν προς τα πίσω, από το τελευταίο επίπεδο στο πρώτο. Αυτός ακριβώς είναι και ο λόγος που ο αλγόριθμος ξεκινά από το τελευταίο επίπεδο και εκεί οφείλεται και το όνομα της μεθόδου.

## Κεφάλαιο 5

# Ανίχνευση αγγειοδυσπλασιών

Έχει ήδη αναφερθεί σε προηγούμενο κεφάλαιο πως η εξέταση του βίντεο της ενδοσκοπικής κάψουλας αποτελεί μια χρονοβόρα και ανιαρή διαδικασία. Αν αναλογιστεί κανείς πως ο χρόνος που απαιτείται για να περάσει η κάψουλα μέσα από το σύνολο του λεπτού εντέρου μπορεί να φτάσει και τις 5 ώρες [44], εύκολα θα καταλάβει πως για να γίνει μια διάγνωση θα είναι απαραίτητη η εξέταση χιλιάδων εικόνων (έως και 70000, για την ακρίβεια), αναλόγως του frame rate της κάψουλας. Συνδυαζόμενο με το γεγονός ότι μια αγγειοδυσπλασία μπορεί να εμφανίζεται μόνο για μερικά ή και για ένα μοναδικό καρέ [45], η ανάγκη προσεκτικής εξέτασης και αμέριστη προσοχής σε αυτή κρίνεται επιτακτική. Για αυτούς τους λόγους, έχουν γίνει πολλές προσπάθειες για ανάπτυξη τεχνολογίας ικανής να αναλύσει τις εικόνες των βίντεο και να αναγνωρίσει αίμα ή άλλες ερυθρόμορφες ανωμαλίες.

### 5.1 State-of-the-art μέθοδοι

Υπάρχουν δύο είδη λογισμικού τα οποία βοηθούν στην διάγνωση ενός CE βίντεο<sup>1</sup>. Στο πρώτο από αυτά ανήκουν λογισμικά τα οποία βελτιώνουν την απεικόνιση και βοηθούν στην αναγνώριση παθολογιών αφού τις κάνουν ευκολότερα διακριτές. Παραδείγματα τέτοιου λογισμικού και τεχνικών είναι το FICE και το blue mode [46]. Ο δεύτερος τύπος είναι λογισμικό το οποίο επιλέγει καρέ από ένα CE βίντεο για επιπλέον εξέταση. Το suspected blood indicator (SBI) εργαλείο, το οποίο αναπτύχθηκε από την Medtronic<sup>®</sup> ως μέρος του λογισμικού ανάγνωσης CE βίντεο RAPID<sup>®</sup>, αποτελεί την state-of-the art μέθοδο σήμερα. Είναι ικανό να ανιχνεύει καρέ με pixels κόκκινης απόχρωσης τα οποία θα μπορούσαν να είναι εστίες πιθανής αιμορραγίας στον γαστρεντερικό σωλήνα. Παρόλα αυτά, η κλινική χρησιμότητα του εργαλείου είναι αμφισβητήσιμη, κυρίως λόγω της περιορισμένης επιτυχίας του στην αναγνώριση, κυρίως, αγγειοδυσπλασιών.

<sup>1</sup>CE Video: Capsule Endoscopy Video, βίντεο ενδοσκοπικής κάψουλας

### **5.1.1 Ανίχνευση αιμορραγιών**

Πρόσφατες έρευνες επάνω στην ακρίβεια του SBI (Han et al. 2018), έδειξαν 100% ευαισθησία στην ανίχνευση των αιμορραγιών. Ταυτόχρονα, όμως, έδειξαν και την ύπαρξη πολλών ψευδώς-θετικών σημείων. Περίπου οι μισές από τις περιπτώσεις ασθενών με ενεργό αιμορραγία παρουσίαζαν ψευδώς θετικές διαγνώσεις αιμορραγίας σε διάφορα σημεία του εντέρου [47]. Οι πλειοψηφία αυτών φέρεται να οφείλεται σε φυσαλίδες οι οποίες ανακλούν φως το οποίο φαίνεται να έχει κόκκινη απόχρωση. Παρόμοια αποτελέσματα έδειξαν και το 2017 ο Yung et al. Πιο συγκεκριμένα, το SBI έδειξε υψηλή ευαισθησία (98.8%) αλλά χαμηλή ειδικότητα (64.6%) [48].

### **5.1.2 Ανίχνευση αγγειοδυσπλασιών**

Αντίστοιχα, για την ανίχνευση αγγειοδυσπλασιών, το λογισμικό SBI εμφανίζει κακή με μέτρια ευαισθησία και ειδικότητα (55.5% και 57.8% αντίστοιχα). Ειδικότερα, σε βίντεο προερχόμενα από την κάψουλα πρώτης γενιάς, τα αποτελέσματα ήταν ακόμη χαμηλότερα (40.3% και 50.5%) [48]. Ως συμπέρασμα, υπάρχουν ακόμη μεγάλοι περιορισμοί στο δρόμο για την απολύτως αυτόνομη ανίχνευση αγγειοδυσπλασιών.

### **5.1.3 Χρόνος εξέτασης**

Σε έρευνα του Omori et. al (2018) [11] ο χρόνος εξέτασης του συνόλου του λεπτού εντέρου με χρήση του συστήματος PillCam SB3 ήταν, κατά μέσο όρο, 11.1 λεπτά για ειδικούς στο πεδίο, ενώ 23.7 λεπτά για αρχάριους.

## **5.2 Άλλες έρευνες και μέθοδοι εντοπισμού παθολογιών**

Αν και εμπορικά το λογισμικό SBI είναι αυτό που χρησιμοποιείται ευρέως, πρόσφατα έχουν γίνει ποικίλες, ανεξάρτητες έρευνες για την ανίχνευση και τη διάγνωση αιμορραγιών και αγγειοδυσπλασιών.

Το 2014, οι Ιακωβίδης και Κουλαουζίδης εξέτασαν την ανίχνευση ανωμαλιών βάσει της διαφοράς του χρώματός τους από τον υπόλοιπο, φυσιολογικό βλεννογόνο [49]. Για κάθε σημείο ενδιαφέροντος πραγματοποιήθηκε εξαγωγή χαρακτηριστικών: η τιμή κάθε χρωματικής συνιστώσας, και οι ελάχιστες και μέγιστες τιμές κάθε χρωματικής συνιστώσας μέσα από μια τετράγωνη γειτονιά pixels. Η επιτυχία της μεθόδου τους διέφερε σημαντικά ανάμεσα σε διαφορετικούς τύπους αγγειοδυσπλασιών (P0, P1 και P2), και κυμαινόταν μεταξύ 69.9% με 97.5%.

Το 2015, οι ίδιοι εξέτασαν την ανίχνευση αίματος σε CE εικόνες βάσει superpixels [50] πετυχαίνοντας ακρίβεια 86%, ευαισθησία 88% και ειδικότητα 84%. Η μέθοδος των superpixels όμως αποτελεί μια εξαιρετικά αργή μέθοδο, η οποία, με μια μικρή δοκιμή, χρειάζεται πάνω από 0.5 δευτερόλεπτα για την εκτέλεσή της. Κάτι τέτοιο την καθιστά απαγορευτική για τη χρησιμοποίησή της, αφού είναι πιο αργή από τη real-time λήψη του βίντεο.

Το 2018, οι Deeba et al. δοκίμασαν μια τεχνική για την ανίχνευση αγγειοδυσπλασιών βάσει της χρωματικής τους ιδιαιτερότητας [51]. Τα αποτελέσματά τους διέφεραν σημαντικά για κάθε βίντεο, με τελικό αποτέλεσμα ευαισθησίας 94.44% και ειδικότητας 83.92%.

Όλες οι παραπάνω έρευνες, αν και πέτυχαν μια αρκετά υψηλή επιτυχία δεν έλαβαν υπόψιν τους μια πληθώρα σημαντικών χαρακτηριστικών. Για παράδειγμα, το γεγονός ότι η ανίχνευση πραγματοποιείται μόνο βάσει του χαρακτηριστικού χρώματος των αγγειοδυσπλασιών σημαίνει ότι δεν λαμβάνεται υπόψιν το σχήμα. Εγείρονται λοιπόν ερωτήματα σχετικά με το πώς τα μοντέλα αυτά θα αντιμετωπίζουν περιοχές με αυξημένη, μεν, αιματική ροή όπου, όμως, δεν υπάρχουν αγγειοδυσπλασίες. Τέτοιες περιοχές, μπορεί να είναι, για παράδειγμα, περιοχές με εμφανή αιμοφόρα αγγεία. Τέτοιες έρευνες συνήθως διαθέτουν ένα περιορισμένο σύνολο δεδομένων (dataset) και συχνά τα ευρήματά τους διαφέρουν ανάμεσα σε διαφορετικά datasets [50]. Θα ήταν λοιπόν επιθυμητό να λαμβάνεται υπόψιν και το σχήμα, και το χρώμα των αγγειοδυσπλασιών.

Άλλο χαρακτηριστικό είναι ο χρόνος υπολογισμού. Η υπολογιστική αποδοτικότητα πρέπει να λαμβάνεται υπόψιν ως μια σημαντική παράμετρος σε τέτοιες μεθόδους, κυρίως λόγω του μεγέθους των βίντεο σε συνδυασμό με τους περιορισμένους υπολογιστικούς πόρους τόσο του ερευνητή όσο και, αργότερα, του προσωπικού το οποίο θα καλείται να χρησιμοποιήσει την νέα τεχνολογία για τη διάγνωση του CE βίντεο [52].

Τέλος, ακόμη ένα σημαντικό χαρακτηριστικό είναι η ευαισθησία των μεθόδων σε παθολογίες με ιδιαίτερα μικρό μέγεθος [52]. Λογισμικά τα οποία λαμβάνουν υπόψιν αποκλειστικά τη χρωματική απόχρωση μέσα σε ένα παράθυρο συγκεκριμένου μεγέθους είναι πολύ πιθανό να εξαρτώνται υπερβολικά από το μέγεθος των ανωμαλιών αυτών στην εικόνα.

Είναι λοιπόν επιθυμητό να αναπτυχθεί μια μέθοδος ανίχνευσης αγγειοδυσπλασιών η οποία να ανταποκρίνεται σε όλες τις παραπάνω απαιτήσεις, πετυχαίνοντας ταυτόχρονα ένα πολύ μεγάλο ποσοστό ευαισθησίας και ένα μεγάλο ποσοστό ειδικότητας.

Στη συνέχεια θα εξηγηθεί αναλυτικά η μέθοδος της αυτόματης διάγνωσης. Ο αναγνώστης θα παρατηρήσει ότι δίνεται έμφαση στην ταχύτητα εκτέλεσης του κάθε βήματος αφού ο σχετικά γρήγορος υπολογισμός χωρίς την ανάγκη πολλών υπολογιστικών πόρων αποτελεί ένα από τα επιθυμητά χαρακτηριστικά της μεθόδου. Ο χρόνος εκτέλεσης κάθε βήματος χρονομετρήθηκε σε ένα απλό laptop, με επεξεργαστή 2.3 GHz Intel Core i5 και μνήμη 8 GB 2133 MHz. Δεν χρησιμοποιήθηκε κάρτα γραφικών για την επιτάχυνση του υπολογισμού των δεδομένων.

### 5.3 Περιοχές ενδιαφέροντος

Ένα πρώτο βήμα στην ανίχνευση των αγγειοδυσπλασιών είναι η επιλογή περιοχών ενδιαφέροντος, δηλαδή περιοχών όπου συναντάται αυξημένη αιματική ροή, ιδίως όταν αυτή εμφανίζει μια ελλειψοειδή γεωμετρία. Είναι επιθυμητό ο αριθμός των περιοχών αυτών να περιορισθεί στον ελάχιστο δυνατό αριθμό, κρατώντας όμως ταυτόχρονα την ευαισθησία σε αγγειοδυσπλασίες υψηλή. Το πλεονέκτημα αυτού είναι ο πολύ μικρότερος αριθμός σημείων που θα πρέπει να ελεγχθούν στην επόμενη φάση του προγράμματος. Έρευνα η οποία χρησιμοποιούσε μέθοδο παρόμοια με αυτή που πρόκειται να αναλυθεί ανίχνευε και αξιολογούσε γύρω στα 14000 σημεία [49]. Με την παρακάτω μέθοδο απομονώνονται μόνο 3 με 5 σημεία ενδιαφέροντος, πετυχαίνοντας πολύ υψηλή ευαισθησία.



Σχήμα 5.1: Εικόνα από CE βίντεο με οριακά εμφανή αγγειοδυσπλασία

#### 5.3.1 Χώρος CIE Lab

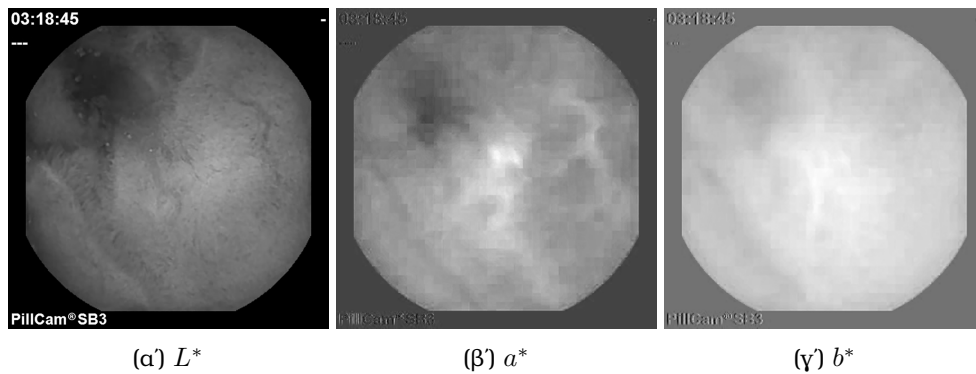
Ο τυπικός χρωματικός χώρος ενός βίντεο είναι ο RGB. Πρόκειται για ένα γραμμικό χώρο ο οποίος παριστάνει ένα χρώμα ως γραμμικό συνδυασμό τριών πρωταρχικών χρωμάτων. Ο χώρος όμως αυτός δεν αντιστοιχεί καλά στην ανθρώπινη αντίληψη χρώματος, αφού οι συντεταγμένες ενός χρώματος στον χώρο αυτό δεν περιέχουν ιδιότητες όπως είναι η απόχρωση, η φωτεινότητα και ο κορεσμός. Εύκολα αντιλαμβάνεται κανείς ότι από αυτές τις ιδιότητες, η απόχρωση αποτελεί την πιο σημαντική καθώς καλούμαστε να αναγνωρίσουμε περιοχές οι οποίες ξεχωρίζουν κυρίως λόγω της ιδιαίτερα κόκκινης απόχρωσής τους σε σύγκριση με το φυσιολογικό βλενογόνο του λεπτού εντέρου. Επιπλέον, θέλοντας να προσεγγίσουμε την ανθρώπινη αντίληψη των διαφορών χρώματος είναι επιθυμητό να χρησιμοποιήσουμε έναν ομοιόμορφο ιστροπικό χρωματικό χώρο, ώστε η απόσταση δύο σημείων στο χώρο αυτό



να αποτελεί μια πιστή ένδειξη για τη διαφορά δύο χρωμάτων από έναν ανθρώπινο παρατηρητή.

Ένας ιδανικός χώρος για αυτή τη δουλειά είναι ο CIE Lab χώρος. Υπενθυμίζεται ότι συνιστώσες του χώρου προκύπτουν από τον εξής μη γραμμικό μετασχηματισμό του χώρου XYZ:

$$\begin{aligned} L^* &= 116 \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 \\ a^* &= 500 \left[ \left( \frac{X}{X_n} \right)^{\frac{1}{3}} - \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} \right] \\ b^* &= 200 \left[ \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left( \frac{Z}{Z_n} \right)^{\frac{1}{3}} \right] \end{aligned} \quad (5.1)$$



Σχήμα 5.2: Τα 3 κανάλια της Lab εικόνας.

Χρόνος που απαιτείται για τη μετατροπή της εικόνας από RGB σε CIE Lab  $\simeq 1$ ms.

### 5.3.2 K-means clustering

Στο επόμενο στάδιο είναι επιθυμητή η απομόνωση των περιοχών εντονότερης κόκκινης απόχρωσης από των υπολοίπων. Η βασική ιδέα πίσω από αυτό το βήμα είναι ότι τα κυρίαρχα χρώματα σε ένα CE βίντεο συνήθως ανήκουν και μπορούν να ομαδοποιηθούν σε τουλάχιστον τρεις κατηγορίες: τα πράσινα, τα κίτρινα, και τα κόκκινα.

Η ομαδοποίηση των δεδομένων βάσει μιας πληροφορίας αποκαλείται, με την αγγλική ονομασία, cluster analysis. Ο στόχος είναι όλα τα μέλη μιας ομάδας να εμφανίζουν παρόμοια χαρακτηριστικά, ή να συσχετίζονται με κάποιο τρόπο μεταξύ τους, ενώ ταυτόχρονα να είναι διαφορετικά από μέλη σε άλλες ομάδες. Όσο μεγαλύτερη η ομοιογένεια μέσα σε μια ομάδα, τόσο μεγαλύτερη η διαφορά μεταξύ των ομάδων, και τόσο καλύτερη είναι η ομαδοποίηση, ή αλλιώς, το clustering.

Ας θεωρήσουμε  $K$  σε αριθμό αρχικά κέντρα βάρους, όπου  $K$  μια παράμετρος η οποία επιλέγεται ελεύθερα και αυθαίρετα, και ισούται με τον αριθμό των επιθυμητών ομάδων στις οποίες θα χωρίσουμε τα δεδομένα μας. Ας θεωρήσουμε ότι τα δεδομένα μπορούν να απεικονισθούν στις 2 διαστάσεις σε ένα σύστημα καρτεσιανών συντεταγμένων. Κάθε δεδομένο απεικονίζεται ως ένα σημείο κάπου στο διδιάστατο χώρο και

ανατίθεται στο πλησιέστερο κέντρο βάρους. Το σύνολο των δεδομένων το οποίο έχει ανατεθεί σε ένα κοινό κέντρο βάρους αποτελεί μία ομάδα, ή αλλιώς, cluster.

Για να ανατεθεί ένα σημείο στο πλησιέστερο κέντρο βάρους χρειαζόμαστε ένα τρόπο μέτρησης της απόστασης ο οποίος θα ποσοτικοποιεί τον όρο «πλησιέστερο» για τα δεδομένα μας. Η Ευκλίδεια ( $L^2$ ) απόσταση είναι αυτή που χρησιμοποιείται συχνότερα για σημεία δεδομένων σε Ευκλίδειο χώρο και ορίζεται για δύο  $n$ -διάστατα σημεία  $\mathbf{p}$  και  $\mathbf{q}$  ως:

$$\begin{aligned} \text{dist}(\mathbf{p}, \mathbf{q}) &= \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \\ &= \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \end{aligned} \quad (5.2)$$

Τώρα θα ορίσουμε το άθροισμα τετραγωνισμένου σφάλματος ( $\text{SSE}^2$ ), γνωστό και ως «διασκορπισμός» (scatter). Ο διασκορπισμός αποτελεί ένα μέτρο της ποιότητας της ομαδοποίησης και υπολογίζεται ως το άθροισμα του τετραγώνου του σφάλματος για κάθε σημείο δεδομένων, δηλαδή η ευκλίδεια απόστασή του από το πλησιέστερο κέντρο βάρους.

$$\text{SSE} = \sum_{i=1}^K \sum_{\mathbf{x} \in C_i} \text{dist}(\mathbf{c}_i, \mathbf{x})^2 \quad (5.3)$$

όπου  $\mathbf{c}_i$  το πλησιέστερο κέντρο βάρους για το σημείο  $\mathbf{x}$ . Δεδομένων δύο διαφορετικών ομαδοποιήσεων επάνω στο ίδιο σύνολο δεδομένων, προτιμούμε αυτό με το μικρότερο SSE καθώς αυτό σημαίνει ότι τα κέντρα βάρους σε αυτή την ομαδοποίηση είναι πιο αντιπροσωπευτικά για τα σημεία τους. Αποδεικνύεται πως το κέντρο βάρους το οποίο ελαχιστοποιεί το SSE της κάθε ομάδας είναι ο μέσος όρος:

$$\mathbf{c}_i = \frac{1}{m_i} \sum_{\mathbf{x} \in C_i} \mathbf{x} \quad (5.4)$$

Μπορούμε τώρα να ορίσουμε το γενικό αλγόριθμο του K-means clustering. Ως πρώτο βήμα επιλέγουμε  $K$  σημεία τα οποία είναι και τα αρχικά κέντρα βάρους. Στη συνέχεια δημιουργούμε  $K$  ομάδες αναθέτοντας κάθε σημείο στο πλησιέστερο κέντρο βάρους και έπειτα επαναυπολογίζουμε τα κέντρα βάρους κάθε ομάδας ελαχιστοποιώντας το SSE. Επαναλαμβάνουμε τα δύο τελευταία βήματα μέχρι να σταματήσουν να αλλάζουν τα κέντρα βάρους. Ο υπολογισμός αυτός μπορεί να μας οδηγήσει σε ένα τοπικό ελάχιστο για το SSE.

Για την εφαρμογή του K-means clustering στην CIE Lab εικόνα, αρχικά εξάγουμε τις  $a^*$  και  $b^*$  συνιστώσες. Υπενθυμίζεται πως αυτές οι συνιστώσες παριστάνουν τις διαφορές μεταξύ ανταγωνιστικών χρωμάτων. Για  $a^* > 0$  έχουμε κόκκινο χρώμα, για  $a^* < 0$  πράσινο, για  $b^* > 0$  έχουμε κίτρινο και για  $b^* < 0$  έχουμε μπλε.

---

<sup>2</sup>Sum of Squared Error

Αν η εικόνα προς ανάλυση έχει μέγεθος  $n \times m$  τότε εξάγουμε δύο πίνακες, έστω  $A_{n \times m}$  και  $B_{n \times m}$  για το  $a^*$  και  $b^*$  κανάλι αντίστοιχα. Αυτούς τους πίνακες είναι δυνατόν να τους αναδιατάξουμε με τέτοιο τρόπο ώστε να έχουμε  $A_{nm \times 1}$  και  $B_{nm \times 1}$ . Αλλάξαμε δηλαδή το μέγεθός τους και τους ορίσαμε σε 1 διάσταση χωρίς απώλεια όμως πληροφορίας. Θεωρούμε ένα καρτεσιανό σύστημα συντεταγμένων όπου ο  $x$  άξονας αντιστοιχεί στις τιμές του πίνακα  $A$ , και ο  $y$  άξονας στις τιμές του πίνακα  $B$ . Μπορούμε να ομαδοποιήσουμε τις χρωματικές διαφορές με τη μέθοδο του K-means clustering, ακριβώς όπως αναφέρθηκε παραπάνω.

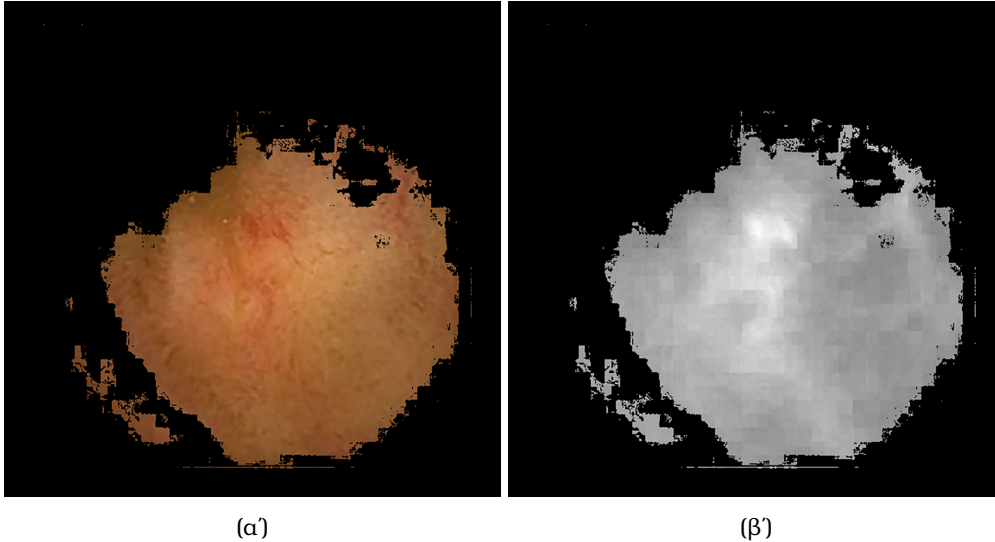
Καθώς είναι επιθυμητή η ομαδοποίηση των δεδομένων ανάμεσα από τρεις κατηγορίες, επιλέγεται  $K = 3$ . Τα κέντρα βάρους αρχικά επιλέγονται τυχαία. Ο αλγόριθμος υπολογίζει σε κάθε επανάληψη τη θέση των κέντρων έτσι ώστε να ελαχιστοποιείται κάθε φορά το SSE. Οι επαναλήψεις σταματούν όταν φτάσουν τις 10 σε αριθμό, ή όταν επιτευχθεί ακρίβεια  $\epsilon = 1$ , δηλαδή όταν τα κέντρα βάρους μετακινηθούν λιγότερο από 1.

Στη συνέχεια, θέλουμε να εφαρμόσουμε μια μάσκα στην εικόνα, έτσι ώστε οι συνιστώσες όλων των pixels, εκτός από αυτών που ανήκουν στην ομάδα της κόκκινης απόχρωσης, να πάρουν την τιμή 0 (να γίνουν, δηλαδή, μαύρα). Μια σημαντική λεπτομέρεια είναι ότι καθώς τα κέντρα βάρους για την κάθε ομάδα επιλέχθηκαν τυχαία, δεν γνωρίζουμε αρχικά ποιά ομάδα αντιστοιχεί στα pixels κόκκινης απόχρωσης. Η λύση σε αυτό το πρόβλημα όμως είναι σχετικά απλή. Γνωρίζοντας τη θέση κάθε κέντρου βάρους, η ομάδα των pixels κόκκινης απόχρωσης θα έχει το κέντρο βάρους με τη μεγαλύτερη τιμή στον  $x$  άξονα. Ο λόγος είναι προφανής, αν θυμηθεί κανείς ότι για την εκτέλεση της K-means clustering μεθόδου, αναθέσαμε στον άξονα των  $x$  τις τιμές της  $a^*$  χρωματικής συνιστώσας, και στον  $y$  τις τιμές της  $b^*$  χρωματικής συνιστώσας. Επομένως, η τιμή του  $x$  αντιστοιχεί στην τιμή της  $a^*$  συνιστώσας, και συνεπώς, όσο μεγαλύτερη αυτή η τιμή τόσο περισσότερο κόκκινη η απόχρωση. Η εφαρμογή της μάσκας είναι επίσης απλή. Δημιουργούμε μια νέα εικόνα  $M_{n \times m}$  την οποία αντιστοιχίζουμε με την αρχική με τον εξής τρόπο: τα pixels τα οποία ανήκουν στην ομάδα της κόκκινης απόχρωσης λαμβάνουν την τιμή 1, ενώ όλα τα υπόλοιπα την τιμή 0. Τελικά, η εφαρμογή της μάσκας γίνεται ως εξής:

$$G = A \odot M \quad (5.5)$$

όπου  $A_{n \times m}$  η  $a^*$  συνιστώσα της αρχικής εικόνας,  $G_{n \times m}$  η προκύπτουσα εικόνα και  $\odot$  τελεστής ο οποίος δηλώνει πολλαπλασιασμό στοιχείο-με-στοιχείο μεταξύ πινάκων.

Μια σημαντική λεπτομέρεια αποτελεί το γεγονός ότι η μάσκα εφαρμόστηκε μόνο στο  $a^*$  κανάλι της εικόνας, και όχι σε ολόκληρη την εικόνα συνολικά. Ο λόγος που έγινε αυτό είναι επειδή πλέον, για την ανίχνευση των αγγειοδυσπλασιών, μας απασχολεί μόνο το  $a^*$  κανάλι λόγω του ότι ποσοτικοποιεί τη διαφορά του πράσινου από το κόκκινο. Τόσο η φωτεινότητα,  $L$ , όσο και η διαφορά του κίτρινου από το μπλε,  $b^*$ , αποτελούν, πλέον, άχρηστες πληροφορίες.



Σχήμα 5.3: Η αρχική εικόνα έπειτα από εφαρμογή της μάσκας (α') και η ίδια μάσκα εφαρμοσμένη στο  $a^*$  κανάλι της εικόνας (β')

Τέλος, η εικόνα  $G$  κανονικοποιείται μεταξύ του 0 και του 255. Χρόνος που απαιτείται για την εφαρμογή του K-means clustering  $\simeq 60\text{ms}$ . Χρόνος που απαιτείται για την εφαρμογή της μάσκας  $\simeq 1\text{ms}$

### 5.3.3 Contrast, Gaussian filtering

Έχοντας πλέον απομονώσει τις εντονότερα κόκκινες περιοχές, θέλουμε να συνεχίσουμε την τροποποίηση της εικόνας ώστε οι κόκκινες περιοχές να ξεχωρίζουν ακόμη περισσότερο. Για να γίνει αυτό, επιθυμούμε να «μαυρίσουμε» τα γκριζα pixels, δηλαδή περιοχές οι οποίες δεν ξεχωρίζουν για την ερυθρόμορφη χρωματική τους απόχρωση, ενώ ταυτόχρονα να ενισχύσουμε ακόμη περισσότερο pixels τα οποία ξεχώριζαν από πριν. Με πιο απλά λόγια, θέλουμε να μεταβάλλουμε την αντίθεση της εικόνας ώστε να «κάψουμε» pixels τα οποία δεν μας ενδιαφέρουν και να γίνουν ένα με το φόντο, ενώ ταυτόχρονα ενισχύουμε τη διαφορά των ιδιαίτερως λευκών pixels (δηλαδή κόκκινης απόχρωσης αφού αναφερόμαστε στο  $a^*$  κανάλι). Για να γίνει αυτό, θέτουμε σε κορεσμό όλα τα pixels έξω από ένα ανώτερο και ένα κατώτερο όριο, ενώ τα υπόλοιπα τα αντιστοιχίζουμε γραμμικά σε ένα νέο σύνολο τιμών, το οποίο στην προκειμένη περίπτωση είναι το σύνολο τιμών της γκριζας στάθμης της εικόνας, 0 έως 255. Θεωρώντας ως κατώτερο όριο το  $t_l$  και ανώτερο το  $t_u$  έχουμε:

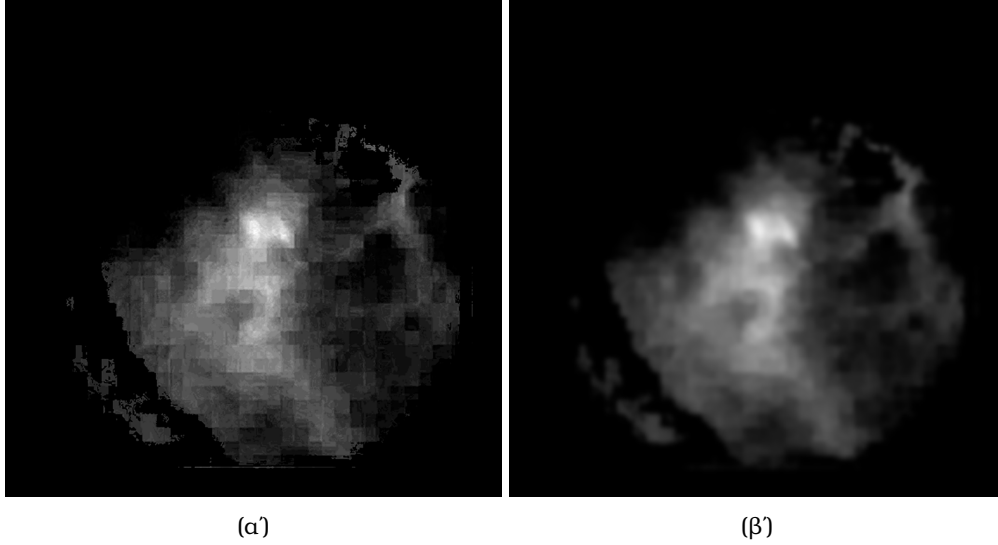
$$G(i, j) = (I(i, j) - t_l) \cdot \frac{255}{t_u - t_l} \quad (5.6)$$

Στη συνέχεια θέτουμε:

$$G(i, j) = \begin{cases} 0, & \text{εάν } G(i, j) < 0 \\ 255, & \text{εάν } G(i, j) > 255 \\ \text{σταθ.}, & \text{αλλού} \end{cases} \quad (5.7)$$

Για τις ανάγκες της μεθόδου χρησιμοποιήθηκαν οι τιμές  $t_l = 140$ ,  $t_u = 255$ .

Άμεσα γίνεται αντιληπτό πως η εικόνα μας πλέον εμφανίζει θόρυβο και τεχνουργήματα (artifacts). Είναι επιθυμητό αυτά να μειωθούν ή και να εξαφανισθούν εντελώς, και για αυτό το λόγο εφαρμόζεται ένα γκαουσιανό φίλτρο στην εικόνα μας. Είναι σημαντικό η σταθερά  $\sigma$  να επιλεγεί κατάλληλα ώστε να έχουμε την απαραίτητη εξομάλυνση στην εικόνα μας, χωρίς όμως να καταστρέφουμε πληροφορία για ιδιαίτερα μικρές αγγειοδυσπλασίες. Πειραματικά, βρέθηκε ότι μια καλή τιμή είναι  $\sigma = 3$ . Χρόνος που απαιτείται για τη μεταβολή του contrast και του γκαουσιανού φιλτραρίσματος  $\simeq 5\text{ms}$



Σχήμα 5.4: Η εικόνα ύστερα από μεταβολή της αντίθεσης (α') και ύστερα από εφαρμογή του γκαουσιανού φίλτρου εξομάλυνσης (β')

### 5.3.4 Speeded-Up Robust Feature Detection

Έχοντας πλέον τροποποιήσει την εικόνα κατάλληλα ώστε να ξεχωρίζουν όσο το δυνατόν περισσότερο οι περιοχές αυξημένης αιματικής ροής, σειρά έχει ο εντοπισμός σημείων ενδιαφέροντος. Λόγω του σχήματός τους, οι αγγειοδυσπλασίες στις εικόνες θα εμφανίζονται ως «blobs», δηλαδή περιοχές με κάποια ομοιογένεια που διαφέρουν σημαντικά από τη γειτονιά τους. Ακριβώς για αυτό το λόγο θα γίνει χρήση ενός γνωστού, αποτελεσματικού και γρήγορου blob detector, του SURF [54].

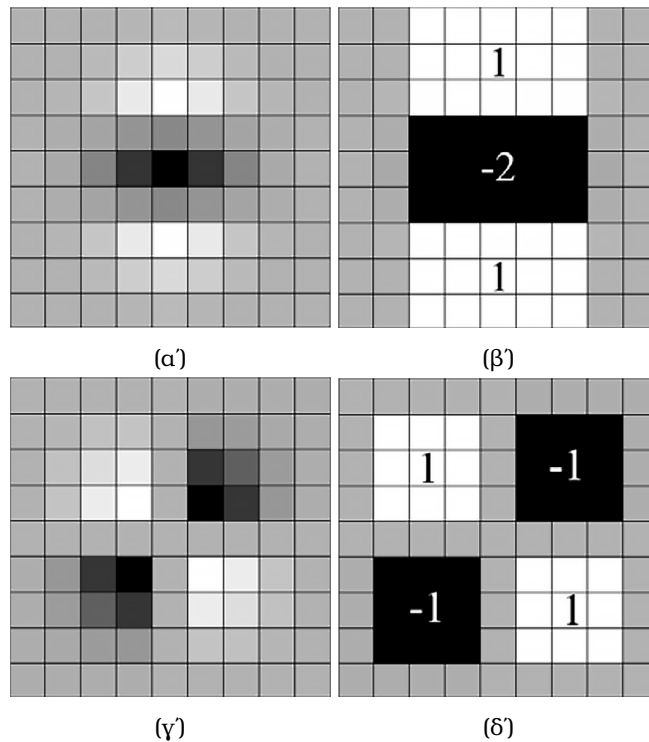
Ο ανιχνευτής αυτός βασίζεται στην Εσσιανή μήτρα, δηλαδή στις μερικές παραγώγους δεύτερης τάξης της εικόνας, λόγω της καλής υπολογιστικής αποδοτικότητας και της ακρίβειάς του. Δοθέντος ενός σημείου  $(x, y)$  σε μια εικόνα  $I$ , ο Εσσιανός (Hessian) πίνακας  $H(x, y, \sigma)$  στο  $(x, y)$  με κλίμακα  $\sigma$  ορίζεται ως εξής:

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix} \quad (5.8)$$

όπου  $L_{xx}(x, y, \sigma) = \frac{\partial^2}{\partial x^2} \{I_\sigma(x, y)\}$ ,  $L_{yy}(x, y, \sigma) = \frac{\partial^2}{\partial y^2} \{I_\sigma(x, y)\}$ ,  $L_{xy}(x, y, \sigma) =$

$\frac{\partial^2}{\partial x \partial y} \{I_\sigma(x, y)\}$ , και  $I_\sigma(x, y) = G_\sigma * I(x, y)$ ,  $G_\sigma$  η γκαουσιανή συνάρτηση.

Καθώς ο υπολογισμός της Hessian για διάφορες κλίμακες  $\sigma$  αντιστοιχεί στη συνέλιξη της εικόνας  $I$  με αυξανόμενου μεγέθους Γκαουσιανά φίλτρα, η μέθοδος είναι ιδιαίτερα απαιτητική από υπολογιστική άποψη. Για την επιτάχυνση αυτής της μεθόδου, τα φίλτρα δεύτερης παραγωγού μπορούν να προσεγγιστούν με τα λεγόμενα *box filters*. Ένα πολύ μεγάλο πλεονέκτημα της προσέγγισης αυτής είναι ότι η συνέλιξη με ένα box filter μπορεί να υπολογιστεί πολύ εύκολα με τη βοήθεια των *ολοκληρωτικών εικόνων*. Άλλο ένα πλεονέκτημα είναι ότι ο υπολογισμός μπορεί να γίνει παράλληλα για πολλαπλές κλίμακες.



Σχήμα 5.5: Οι διακριτοποιημένες δεύτερης τάξης Γκαουσιανές μερικές παράγωγοι μαζί με τις προσεγγίσεις τους με χρήση box filters στην διεύθυνση του  $y$  και του  $xy$  αντίστοιχα. Οι γκρι περιοχές είναι ίσες με το 0.

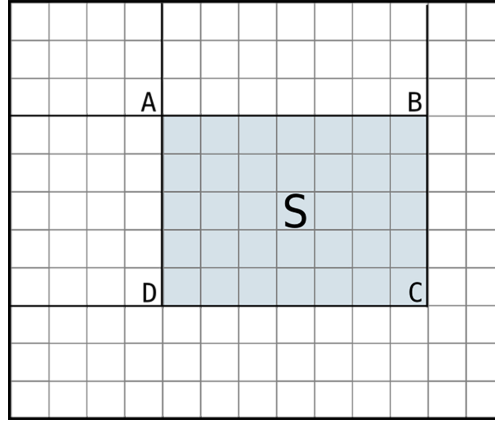
Ας θεωρήσουμε μια εικόνα  $I(x, y)$  μεγέθους  $n \times m$ . Η ολοκληρωτική εικόνα υπολογίζεται ως:

$$S(x, y) = \sum_{i \leq x} \sum_{j \leq y} I(i, j) \quad (5.9)$$

Δηλαδή, κάθε pixel της ολοκληρωτικής εικόνας ισούται με το άθροισμα όλων των στοιχείων της αρχικής εικόνας  $I$  που βρίσκονται πάνω και αριστερά από το σημείο αυτό. Ο υπολογισμός αυτός επιταχύνεται ακόμη περισσότερο εάν παρατηρηθεί ότι κάθε νέο στοιχείο της ολοκληρωτικής εικόνας  $S(i, j)$  μπορεί να υπολογιστεί απευθείας με τη χρήση των γειτονικών του, ήδη υπολογισθέντων, στοιχείων  $S(i, j - 1)$ ,  $S(i - 1, j)$  και  $S(i - 1, j - 1)$  ως εξής:

$$S(i, j) = I(i, j) - S(i - 1, j - 1) + S(i, j - 1) + S(i - 1, j) \quad (5.10)$$

όπου  $i \in [1, n]$  και  $j \in [1, m]$ .



Σχήμα 5.6: Υπολογισμός του ολοκληρωτικού αθροίσματος  $S$

Με αυτόν τον τρόπο, οποιοδήποτε άθροισμα εντός ενός παραθύρου επάνω στην αρχική εικόνα μπορεί να υπολογιστεί άμεσα. Επομένως, οποιοδήποτε φίλτρο βασίζεται σε υπολογισμό αθροισμάτων εντός ορθογώνιων παραθύρων μπορεί να υπολογιστεί πολύ εύκολα και αποδοτικά με την χρήση ολοκληρωτικών εικόνων με πολυπλοκότητα ανεξάρτητη του μεγέθους του φίλτρου.

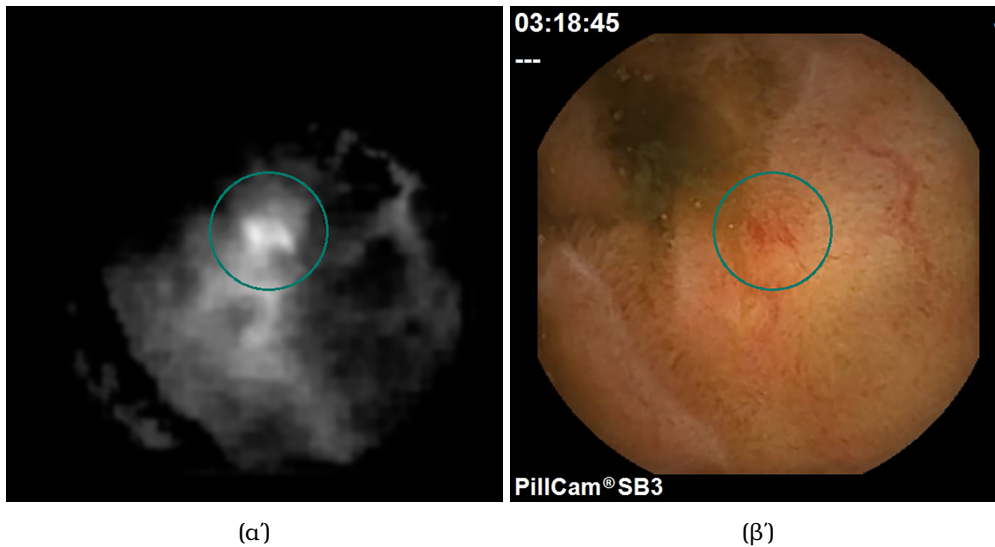
Ας ονομάσουμε τα box filters ως  $D_{xx}$  για την προσέγγιση του  $L_{xx}(x, y, \sigma)$ ,  $D_{yy}$  για το  $L_{yy}(x, y, \sigma)$ , και  $D_{xy}$  για το  $L_{xy}(x, y, \sigma)$ . Αποδεικνύεται [54] ότι η ορίζουσα του Εσσιανού πίνακα μπορεί να υπολογιστεί χρησιμοποιώντας την ολοκληρωτική εικόνα ως εξής:

$$\det(H) \approx D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (5.11)$$

Προκειμένου να εντοπιστούν τα σημεία ενδιαφέροντος στην εικόνα για πολλαπλές κλίμακες, εξετάζονται διαδοχικά εικόνες διαφορετικής, κάθε φορά, κλίμακας με χρήση Γκαουσιανών πυρήνων (kernels). Έτσι, ο χώρος κλίμακας διαιρείται σε επίπεδα και οκτάβες. Η κάθε οκτάβα αντιστοιχεί σε διπλασιασμό του  $\sigma$ , ενώ ταυτόχρονα διαιρείται σε ομοιόμορφα επίπεδα. Για να βρεθεί ένα σημείο ενδιαφέροντος, βρίσκονται τα τοπικά μέγιστα της ορίζουσας του Εσσιανού πίνακα τα οποία εμφανίζονται σε διαδοχικές κλίμακες, και τα οποία είναι μεγαλύτερα από ένα κατώτερο όριο, το threshold (1000 για τις ανάγκες αυτής της μεθόδου).

Με αυτό τον τρόπο ανιχνεύονται σημεία ενδιαφέροντος στην εικόνα, των οποίων είναι πλέον γνωστή η θέση και το μέγεθος μέσω της κλίμακας (σχήμα 5.7). Χρόνος εκτέλεσης του SURF αλγορίθμου  $\simeq 27\text{ms}$ .

Συνολικός χρόνος που απαιτείται για την ανίχνευση των περιοχών ενδιαφέροντος  $\simeq 1 + 60 + 1 + 5 + 27 = 94\text{ms}$ .



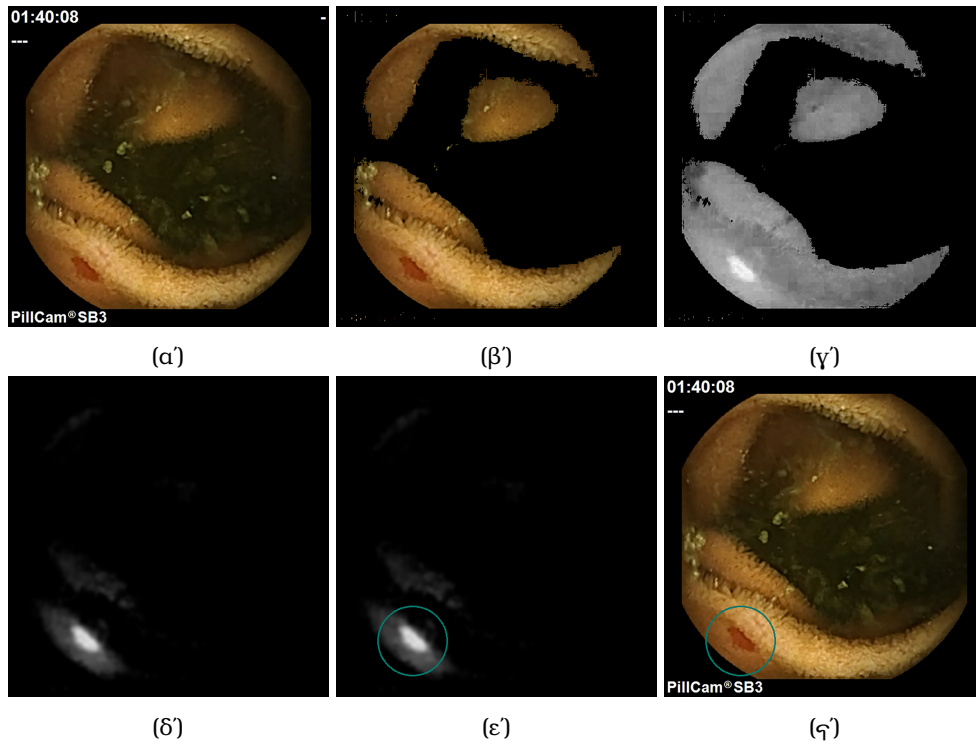
Σχήμα 5.7: Ανίχνευση σημείων ενδιαφέροντος (α') και αντιστοίχιση της περιοχής ενδιαφέροντος στην αρχική εικόνα (β').

## 5.4 CNN κατηγοριοποίηση

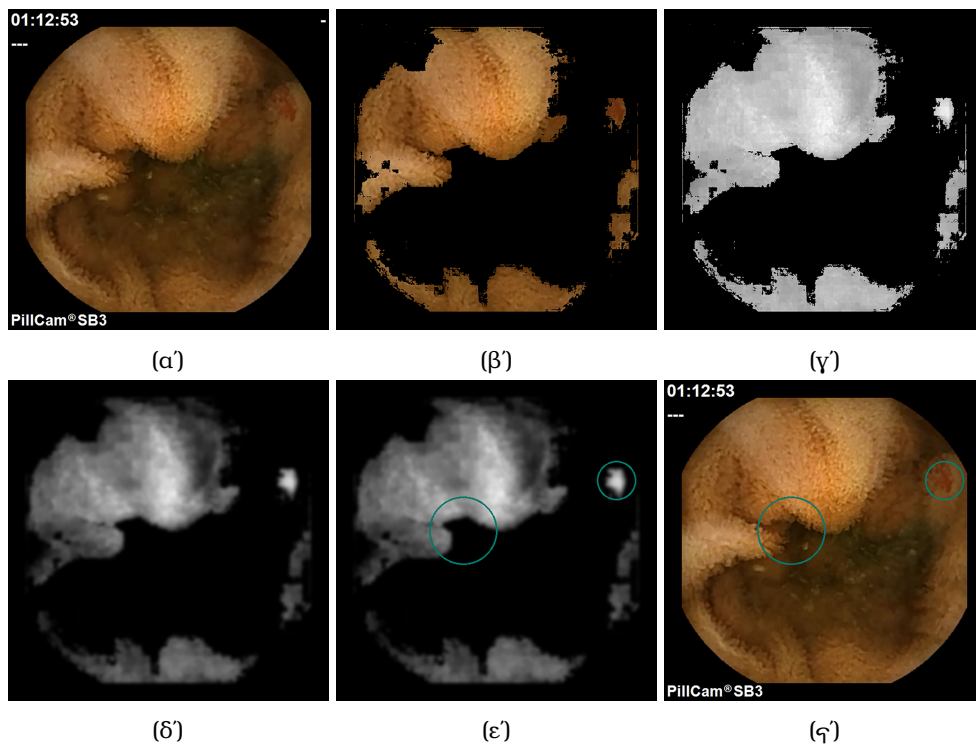
Κατόπιν της εύρεσης σημείων ενδιαφέροντος, είναι αναγκαία η διερεύνησή τους. Ο αλγόριθμος ο οποίος έχει αναλυθεί μέχρι στιγμής είναι ιδιαίτερα αποτελεσματικός στην εύρεση αγγειοδυσπλασιών, ωστόσο πολύ συχνά προτείνει ως περιοχές ενδιαφέροντος και φυσιολογικά σημεία, τα οποία μπορεί να οφείλονται σε αυξημένη αιματική ροή στην περιοχή, μπορεί όμως και να αποτελούν κάτι εντελώς διαφορετικό. Ο λόγος γίνεται αντιληπτός από την επανεξέταση του αλγορίθμου SURF, ο οποίος υπενθυμίζεται πως ανιχνεύει όλα τα σημεία στα οποία εμφανίζει τοπικό μέγιστο η ορίζουσα του Εσσιανού πίνακα σε διαδοχικές κλίμακες. Αυτό μπορεί να συμβαίνει είτε σε λευκά, είτε σε μαύρα σημεία όπως απεικονίζονται στην CIE Lab εικόνα (βλ. σχήμα 5.9).

Είναι αναγκαία η περαιτέρω μείωση των, προτεινόμενων ως παθολογικών, περιοχών. Για το σκοπό αυτό χρησιμοποιείται ένα συνελκτικό νευρωνικό δίκτυο, εκπαιδευμένο έτσι ώστε να αναγνωρίζει και να ξεχωρίζει αγγειοδυσπλασίες από τον φυσιολογικό βλενογόνο, από τις φυσαλίδες και από αιμοφόρα αγγεία τα οποία δεν εμφανίζουν κάποια παθολογία. Ο λόγος που χρησιμοποιείται ένα νευρωνικό δίκτυο είναι λόγω της υψηλής ακρίβειας στην κατηγοριοποίηση εικόνων που μπορούν, πλέον, να επιτύχουν, ιδιαίτερα όταν πρόκειται για λίγες κλάσεις με εικόνες οι οποίες εμφανίζουν σχετικά λίγες ιδιομορφίες. Ένα νευρωνικό δίκτυο μπορεί να λάβει υπόψιν, εκτός από το χρώμα, και το σχήμα των αγγειοδυσπλασιών, αλλά και άλλων ιδιομορφιών. Καλύπτει, λοιπόν, τις ανάγκες οι οποίες έχουν αναφερθεί προηγουμένως.





Σχήμα 5.8: Επιπλέον παράδειγμα ανίχνευσης περιοχής ενδιαφέροντος με εμφανή αγγειοδυσπλασία



Σχήμα 5.9: Το K-means clustering μπορεί να απομονώσει και να καταστήσει τις αγγειοδυσπλασίες ακόμη πιο εμφανείς. Η μέθοδος SURF όμως δεν επιλέγει μόνο αγγειοδυσπλασίες ως περιοχές ενδιαφέροντος.

### 5.4.1 Συνελικτικά νευρωνικά δίκτυα

Τα συνελικτικά νευρωνικά δίκτυα (CNNs / ConvNets<sup>3</sup>) είναι αρκετά όμοια με τα νευρωνικά δίκτυα τα οποία έχουν ήδη επεξηγηθεί. Και τα δύο αποτελούνται από νευρώνες, οι οποίοι στηρίζονται στην εκμάθηση βαρών και πολώσεων. Κάθε νευρώνας δέχεται εισόδους, και υπολογίζει ένα σταθμισμένο άθροισμα το οποίο στη συνέχεια το περνά μέσα από μια μη-γραμμική συνάρτηση. Επίσης, η έξοδος και των δύο αποτελείται από βαθμονόμηση των κλάσεων στις οποίες μπορεί να ανήκει η είσοδος. Η διαφορά τους, όμως, έγκειται στο ότι οι αρχιτεκτονικές των ConvNets αντιμετωπίζουν τις εισόδους ως εικόνες, το οποίο επιτρέπει την εφαρμογή ιδιοτήτων οι οποίες καθιστούν το νευρωνικό δίκτυο πιο αποδοτικό και με λιγότερες παραμέτρους. Με τον όρο *αρχιτεκτονική* εννοείται ο τρόπος κατασκευής του δικτύου, δηλαδή το σχήμα, μέγεθος και αριθμός των επιπέδων.

Υπενθυμίζεται ότι τα κλασικά νευρωνικά δίκτυα δέχονται ένα διάνυσμα εισόδου το οποίο και μετασχηματίζουν μέσα από μια σειρά κρυφών επιπέδων. Κάθε κρυφό επίπεδο αποτελείται από νευρώνες, με τον κάθε νευρώνα να συνδέεται πλήρως μέσω των συνάψεων με όλους τους νευρώνες του προηγούμενου επιπέδου. Κάθε νευρώνας λειτουργεί απολύτως ανεξάρτητα από τους υπόλοιπους του επιπέδου του. Το τελευταίο, πλήρως-συνδεδεμένο, επίπεδο, ονομάζεται επίπεδο εξόδου και αναπαριστά την βαθμονόμηση των κλάσεων.

Αυτά τα κλασικά δίκτυα δεν λειτουργούν πολύ καλά με εικόνες. Ας φανταστούμε μια εικόνα μεγέθους  $256 \times 256 \times 3$ . Πρόκειται για μία εξαιρετικά μικρή εικόνα. Προκειμένου η εικόνα αυτή να δοθεί ως είσοδος σε ένα κλασικό νευρωνικό δίκτυο, χρειάζεται να την μετασχηματίσουμε σε ένα μονοδιάστατο διάνυσμα μεγέθους  $256 \cdot 256 \cdot 3 = 196608$ . Αν δοκιμάσουμε να προσθέσουμε ένα κρυφό επίπεδο με ένα, μόνο, νευρώνα, τότε απευθείας εισάγονται και  $1 \cdot 196608 = 196608$  βάρη. Αν, δε, στο επίπεδο αυτό προστεθεί και δεύτερος νευρώνας, τα βάρη θα γίνουν  $2 \cdot 196608 = 393216$ . Σχεδόν 400.000 βάρη για δύο εξόδους, ή αλλιώς, δύο ενεργοποιήσεις (activations), με ένα μόνο κρυφό επίπεδο. Γρήγορα γίνεται αντιληπτό ότι αυτού του είδους η σύνδεση των επιπέδων συνεπάγεται σε έναν υπέρογκο αριθμό παραμέτρων. Συνέπεια των υπερβολικών παραμέτρων είναι το *overfitting*, ένας εχθρός για τα νευρωνικά δίκτυα ο οποίος μειώνει την αποτελεσματικότητά τους και καθιστά αδύνατη τη γενίκευσή τους.

Τα συνελικτικά νευρωνικά δίκτυα, σε αντίθεση με τα κλασικά, εκμεταλλεύονται το γεγονός ότι η είσοδος αποτελείται από εικόνες. Πιο συγκεκριμένα, τα επίπεδα των ConvNets διαθέτουν νευρώνες διατεταγμένους σε 3 διαστάσεις: πλάτος (width), ύψος (height) και βάθος (depth). Τα κύρια είδη των επιπέδων που χρησιμοποιούνται για το σχηματισμό ενός CNN είναι τρία: το συνελικτικό επίπεδο (convolutional layer), το επίπεδο συγκέντρωσης (pooling layer) και το πλήρως-συνδεδεμένο επίπεδο (fully-connected layer). Στο εξής, τα επίπεδα αυτά θα αναφέρονται με την αγγλική τους ονομασία.

---

<sup>3</sup>Convolutional Neural Networks

## Convolutional Layer

Το convolutional layer αποτελεί το βασικό δομικό στοιχείο ενός συνελκτικού δικτύου. Αν και τα κρυφά επίπεδα των κλασικών δικτύων αποτελούνται από νευρώνες, τα επίπεδα στα ConvNets αποτελούνται από εκπαιδευσιμα φίλτρα. Τα φίλτρα αυτά είναι γενικά μικρών διαστάσεων, έχουν δηλαδή μικρό πλάτος και ύψος, ενώ το βάθος τους ταυτίζεται με αυτό της εικόνας εισόδου. Για παράδειγμα, ένα τυπικό φίλτρο στο πρώτο επίπεδο ενός CNN μπορεί να έχει μέγεθος  $5 \times 5 \times 3$ . Όπως μπορεί να έχει ήδη φανταστεί κανείς, τα φίλτρα αυτά αποτελούν στην ουσία πυρήνες (kernels), με τους οποίους συνελίσσουμε την εικόνα κατά τη διαδικασία του forward pass, υπολογίζοντας το σταθμισμένο άθροισμα για κάθε pixel. Ως τελικό αποτέλεσμα από τη συνέλιξη, προκύπτει ένας διδιάστατος πίνακας, μικρότερης διάστασης από την αρχική εικόνα<sup>4</sup>, ο οποίος περιέχει τις αποκρίσεις του αντίστοιχου φίλτρου. Ο πίνακας αυτός ονομάζεται *activation map*, σε απόλυτη αντιστοιχία με το απλό activation που έχουμε στα κλασικά δίκτυα. Διαισθητικά, αναλόγως την περίπτωση, το δίκτυο μαθαίνει φίλτρα τα οποία αναγνωρίζουν κάποιο οπτικό χαρακτηριστικό, όπως είναι για παράδειγμα κάποια ακμή, ή κάποιο χρώμα, για χαμηλά επίπεδα, και κάποιο πολύπλοκο χαρακτηριστικό, για παράδειγμα ένα κυψελοειδές μοτίβο, σε υψηλότερα επίπεδα. Τα activation maps τα οποία προκύπτουν από κάθε φίλτρο σε ένα κρυφό επίπεδο, τα σιθάζουμε στη διάσταση του βάθους και αποτελούν την έξοδο του κάθε επιπέδου.

Όπως στα κλασικά δίκτυα χρησιμοποιούμε μια μη-γραμμική συνάρτηση για το τελικό αποτέλεσμα της ενεργοποίησης, έτσι και στα συνελκτικά περνάμε την έξοδο κάθε convolutional layer από μια μη-γραμμικότητα. Η μη-γραμμικότητα αποτελεί στην πραγματικότητα ένα επιπλέον επίπεδο το οποίο ακολουθεί έπειτα από κάθε convolutional layer και λέγεται *RELU layer*. Το επίπεδο αυτό εφαρμόζει μια μη-γραμμική συνάρτηση σε κάθε στοιχείο της εξόδου ξεχωριστά και η έξοδος του ταυτίζεται σε μέγεθος με την είσοδο.

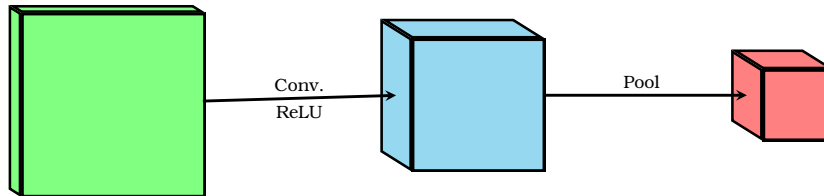
## Pooling Layer

Συχνά, ανάμεσα από διαδοχικά convolutional layers εισάγεται ένα pooling layer. Η λειτουργία συνίσταται στη σταδιακή μείωση του μεγέθους της εξόδου του προηγούμενου επιπέδου, το οποίο έχει ως αποτέλεσμα τη μείωση των παραμέτρων και του πλήθους των υπολογισμών στο δίκτυο. Κάτι τέτοιο βοηθά στον έλεγχο του overfitting. Το pooling layer ενεργεί ανεξάρτητα σε κάθε διάσταση κατά βάθος, και τη μειώνει χωρικά. Η μείωση γίνεται με τη χρήση ενός μικρού παραθύρου, συνήθως  $2 \times 2$ , με stride 2, από το οποίο επιλέγεται η μέγιστη τιμή μέσα σε αυτό. Δηλαδή, από τις 4 τιμές που καλύπτει ένα  $2 \times 2$  παράθυρο, επιλέγεται μόνο 1, οδηγώντας σε μείωση κατά 75% του activation map.

<sup>4</sup>Υπενθυμίζεται ότι κατά τη διαδικασία συνέλιξης χωρίς κάποιο padding, η διάσταση της εικόνας εξόδου είναι μειωμένη σε σχέση με την αρχική. (βλ. σχ. 3.3, σελ. 37)

## Fully-Connected Layer

Συνήθως, στα τελευταία κρυφά επίπεδα εισάγουμε μερικά πλήρως συνδεδεμένα επίπεδα. Τα επίπεδα αυτά είναι ακριβώς ίδια με τα επίπεδα των κλασικών νευρωνικών δικτύων και ο ρόλος τους είναι η κατάταξη της αρχικής εικόνας σε μία από τις τελικές κλάσεις βάσει των τελικών activation maps.



Σχήμα 5.10: Η επίδραση του convolutional και του pooling επιπέδου στην είσοδο. Το convolutional layer μικραίνει το ύψος και το πλάτος της εικόνας, αυξάνει όμως το βάθος αφού όλα τα αποτελέσματα των συνελίξεων σιθάζονται σε αυτή τη διάσταση. Το pooling layer διατηρεί το βάθος, μικραίνει όμως περαιτέρω το ύψος και το πλάτος.

### 5.4.2 Εκπαίδευση

Σε αντίθεση με το κεφάλαιο 4, εδώ δεν θα επεξηγηθεί ο μηχανισμός της εκπαίδευσης, αλλά η διαδικασία που ακολουθήθηκε. Η εκπαίδευση ενός CNN αποτελεί μια διαδικασία η οποία παρουσιάζει προβλήματα, κυρίως λόγω του ότι η εκπαίδευση από το μηδέν απαιτεί ένα πολύ μεγάλο όγκο δεδομένων (training data). Ιδίως σε ιατρικές εφαρμογές, η εύρεση ενός τέτοιου όγκου δεδομένων είναι δύσκολη, τόσο λόγω του κόστους για την εύρεση ιατρικού προσωπικού το οποίο να είναι σε θέση να κατηγοριοποιήσει τα δεδομένα για την εκπαίδευση, όσο και λόγω της σπανιότητας πολλών παθολογιών σε datasets. Ακόμη ένας περιορισμός έγκειται στο ότι η εκπαίδευση από το μηδέν απαιτεί πολλούς υπολογιστικούς πόρους, χωρίς τους οποίους η διαδικασία αυτή απαιτεί ένα απαγορευτικό, σε διάρκεια, χρονικό διάστημα.

Μια πολλά υποσχόμενη εναλλακτική στην εκπαίδευση ενός CNN από το μηδέν, είναι η ακριβής προσαρμογή ενός άλλου δικτύου στα επιθυμητά δεδομένα, γνωστή ως fine-tuning. Πράγματι, ένα μεγάλο πλεονέκτημα των CNNs είναι η δυνατότητα «μεταφοράς» της γνώσης τους. Έρευνες έχουν δείξει ότι η επιτυχία της μεταφοράς της γνώσης (knowledge transfer) εξαρτάται από την απόσταση, ή αλλιώς, από την ανομοιομορφία μεταξύ του συνόλου δεδομένων (dataset) με το οποίο εκπαιδεύεται το νευρωνικό δίκτυο, και του συνόλου δεδομένων στο οποίο θέλουμε να μεταφέρουμε τη γνώση [56].

Όπως έχει ήδη εξηγηθεί, η εκπαίδευση ενός νευρωνικού δικτύου ξεκινά με την αρχικοποίηση του κάθε βάρους που περιέχεται σε αυτό. Η αρχικοποίηση αυτή γίνεται τυχαία μέσα από μια κανονική κατανομή με μηδενική μέση τιμή και μικρή τυπική απόκλιση. Αν ο όγκος του dataset δεν είναι αρκετά μεγάλος, η εκπαίδευση, ή αλλιώς, η ανανέωση του κάθε βάρους, μπορεί να οδηγήσει σε ένα μη-επιθυμητό τοπικό ελάχιστο για τη συνάρτηση κόστους. Εναλλακτικά, τα βάρη μπορούν να

αρχικοποιηθούν χρησιμοποιώντας τα βάρη ενός προ-εκπαιδευμένου δικτύου, με την ίδια αρχιτεκτονική. Η εκπαίδευση ενός δικτύου το οποίο αρχικοποιείται με βάρη τα οποία έχουν προκύψει από την εκπαίδευση ενός άλλου δικτύου ονομάζεται *fine-tuning*, επίσης γνωστό και ως *transfer learning*.

Το *fine-tuning* ξεκινά αντιγράφοντας, δηλαδή μεταφέροντας (transfer), τα βάρη από το προ-εκπαιδευμένο δίκτυο στο δίκτυο το οποίο θέλουμε να εκπαιδεύσουμε. Είναι σημαντικό να αναφερθεί ότι δεν μεταφέρονται όλα τα βάρη. Το τελευταίο, πλήρως συνδεδεμένο, επίπεδο αποτελεί την εξαίρεση, καθώς οι κλάσεις στις οποίες θέλουμε να κατηγοριοποιήσουμε μια εικόνα με το καινούργιο δίκτυο διαφέρουν από τις κλάσεις του προηγούμενου. Επομένως, ο αριθμός των νευρώνων διαφέρει επίσης, και η μεταφορά κάποιων βαρών μεταξύ των δύο δικτύων στο επίπεδο αυτό δεν έχει νόημα. Για παράδειγμα, σε αυτή την εργασία χρησιμοποιήθηκε ένα πλήρως συνδεδεμένο επίπεδο με 4 νευρώνες, οι οποίοι αποτελούσαν την έξοδο και κατέτασαν την εικόνα μεταξύ 4 κλάσεων: φυσιολογικός βλεννογόνας, αγγειοδυσπλασίες, φυμαλίδες, και αιμοφόρα αγγεία.

Ας θεωρήσουμε ένα CNN με  $L$  επίπεδα, όπου τα 3 τελευταία επίπεδα είναι πλήρως-συνδεδεμένα. Έστω ότι  $a_l$  είναι το learning rate του  $l$ -οστού επιπέδου του δικτύου. Μπορούμε να κάνουμε *fine-tune* μόνο το τελευταίο επίπεδο του δικτύου θέτοντας  $a_l = 0$  για  $l \neq L$ . Αυτό ισοδυναμεί με την εκπαίδευση ενός γραμμικού classifier με εισόδους τις αποκρίσεις των προηγούμενων νευρώνων ή φίλτρων από το επίπεδο  $L - 1$ . Ομοίως, μπορούμε να εκπαιδεύσουμε τα 2 τελευταία επίπεδα θέτοντας  $a_l = 0$  για  $l \neq L$  και  $l \neq L - 1$ . Αυτό ισοδυναμεί στην εκπαίδευση ενός νευρωνικού δικτύου με ένα κρυφό επίπεδο με εισόδους τις αποκρίσεις των νευρώνων ή φίλτρων από το επίπεδο  $L - 2$ . Όσο περισσότερα επίπεδα συμπεριλαμβάνουμε στην εκπαίδευση, τόσο περισσότερο προσαρμόζεται το CNN στην επιθυμητή εφαρμογή, ταυτόχρονα όμως απαιτείται ολόένα και περισσότερος όγκος δεδομένων για την αποφυγή του *overfitting*.

Γενικά, τα πρώτα επίπεδα ενός CNN μαθαίνουν να αναγνωρίζουν χαρακτηριστικά χαμηλού επιπέδου, τα οποία βρίσκουν εφαρμογή στις περισσότερες περιπτώσεις όρασης. Αντίθετα, τα τελευταία επίπεδα μαθαίνουν χαρακτηριστικά υψηλού επιπέδου τα οποία είναι εξειδικευμένα στο dataset το οποίο χρησιμοποιήθηκε. Επομένως, το *fine-tuning* των τελευταίων επιπέδων είναι συνήθως αρκετό για την εφαρμογή του *transfer-learning*. Αν, όμως, η ανομοιογένεια των datasets των δύο συνόλων είναι υψηλή, η συμπερίληψη των αρχικών επιπέδων στην εκπαίδευση μπορεί να είναι αναγκαία. Επομένως, μία αποτελεσματική τεχνική είναι να ξεκινήσει κανείς από το τελευταίο επίπεδο και να συμπεριλαμβάνει διαδοχικά επιπλέον επίπεδα μέχρι να επιτευχθεί η επιθυμητή επίδοση.

Εγείρεται, όμως, το ερώτημα του εάν τα προ-εκπαιδευμένα CNNs, μπορούν, έπειτα από επαρκές *fine-tuning*, να χρησιμοποιηθούν στη θέση δικτύων τα οποία έχουν εκπαιδευτεί από το μηδέν, στο πλαίσιο ιατρικών δεδομένων. Πρόσφατες έρευνες έχουν δείξει ότι τα συνελκτικά νευρωνικά δίκτυα τα οποία εκπαιδεύτηκαν με

χρήση της τεχνικής του transfer-learning είναι ιδιαίτερα χρήσιμα και αποτελεσματικά για την ανάλυση ιατρικών εικόνων, με επιδόσεις οι οποίες είναι συγκρίσιμες, εάν όχι καλύτερες, από αντίστοιχες δικτύων τα οποία εκπαιδεύτηκαν από το μηδέν [57].

### 5.4.3 Εφαρμογή στο πρόβλημα αναγνώρισης αγγειοδυσπλασιών

Στην εισαγωγή του κεφαλαίου αυτού αναφέρθηκε ότι κατόπιν εύρεσης των σημείων ενδιαφέροντος στις εικόνες της ενδοσκοπικής κάψουλας, είναι αναγκαία η διερεύνησή τους για την εξακρίβωση του τί απεικονίζεται στις περιοχές αυτές. Τέτοιες περιοχές βρέθηκε ότι μπορεί να ανήκουν σε μια από 4 κατηγορίες:

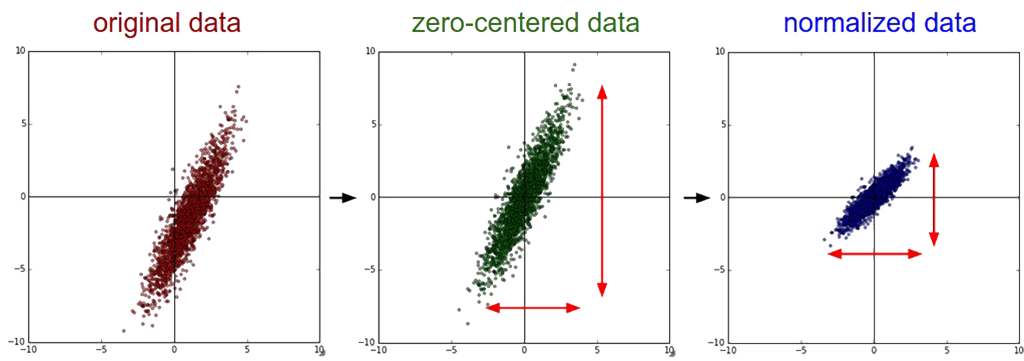
- Φυσιολογικός βλεννογόνος
- Φυσαλίδες
- Αιμοφόρα αγγεία
- Αγγειοδυσπλασίες

Για το σκοπό αυτό δημιουργήθηκε ένα dataset 725, συνολικά, εικόνων, μεγέθους  $227 \times 227$ , οι οποίες κατηγοριοποιήθηκαν σε αντίστοιχους φακέλους ανάλογα της κατηγορίας τους. Πιο συγκεκριμένα, από τις 725 εικόνες, 350 εξ αυτών ανήκαν στην κατηγορία του φυσιολογικού βλεννογόνου, 196 στις φυσαλίδες, 75 στα αιμοφόρα αγγεία και 104 στις αγγειοδυσπλασίες. Στη συνέχεια, οι εικόνες αυτές κανονικοποιούνται προτού δοθούν ως είσοδοι στο νευρωνικό για την εκπαίδευσή του.

Για το πρώτο στάδιο της κανονικοποίησης, η πιο κοινή μέθοδος είναι η *αφαίρεση της μέσης τιμής* (mean subtraction). Πιο συγκεκριμένα, αφαιρείται η μέση τιμή από κάθε στοιχείο των δεδομένων. Δηλαδή, για μια RGB εικόνα, αφαιρείται από κάθε pixel κάθε καναλιού η μέση τιμή που προκύπτει από όλα τα pixels της αντίστοιχης συνιστώσας. Γεωμετρικά, αν φανταστούμε πως τα δεδομένα διατάσσονται σε ένα  $xy$  σύστημα συντεταγμένων, η διαδικασία αυτή τα μετακινεί ώστε το κέντρο τους να βρίσκεται στην αρχή των συντεταγμένων. Για το dataset των αγγειοδυσπλασιών χρησιμοποιήθηκαν οι τιμές 0.485 για το  $R$  κανάλι, 0.456 για το  $G$  κανάλι, και 0.406 για το  $B$  κανάλι. Στο δεύτερο στάδιο, τα στοιχεία κάθε συνιστώσας διαιρούνται με την τυπική απόκλιση του αντίστοιχου καναλιού, αφού όμως αυτό έχει «κεντραριστεί» στο μηδέν. Οι τιμές που χρησιμοποιήθηκαν εδώ είναι οι 0.229, 0.224 και 0.225 για το  $R$ ,  $G$ , και  $B$  κανάλι αντίστοιχα. Αν θεωρήσουμε την εικόνα  $f(x, y, h)$ , η κανονικοποίηση γίνεται σύμφωνα με τη σχέση:

$$f'(x, y, h) = \frac{f(x, y, h) - \text{mean}(h)}{\text{std}(h)} \quad (5.12)$$

Συνεχίζοντας, φορτώνεται ένα προ-εκπαιδευμένο συνελκτικό νευρωνικό δίκτυο, αρχιτεκτονικής ResNet18 [58], στο οποίο και προστίθεται ένα πλήρως συνδεδεμένο επίπεδο με 4 νευρώνες, οι οποίοι αντιστοιχούν στις κατηγορίες που αναφέρθηκαν πιο πάνω. Ως αρχικό learning rate επιλέγεται μια μικρή τιμή, συγκεκριμένα 0.001,

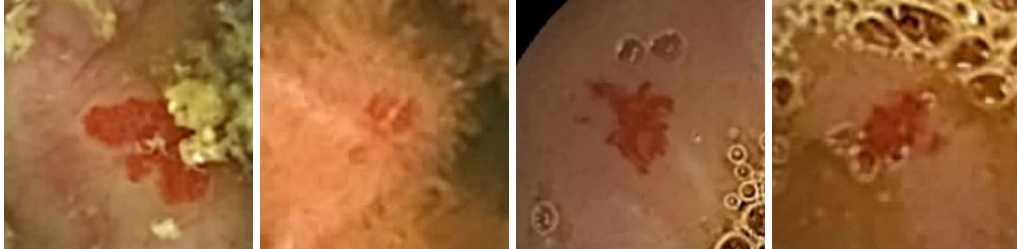


Σχήμα 5.11: Η επίδραση της κανονικοποίησης σε ένα σύνολο δεδομένων.

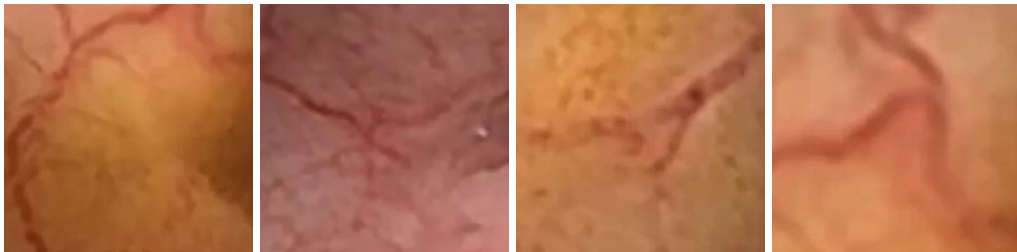
η οποία και μικραίνει καθώς η εκπαίδευση προχωρά. Το νευρωνικό δίκτυο εκπαιδεύεται σύμφωνα με αυτές τις τιμές και δεδομένα.

Χρόνος που απαιτείται για την κατηγοριοποίηση μια εικόνας: 80ms.

Συνολικός μέγιστος χρόνος που απαιτείται για την πλήρη διάγνωση μιας εικόνας (με μέγιστη ανίχνευση 3 περιοχών ενδιαφέροντος)  $94 + 80 \cdot 3 = 334\text{ms}$ . Δηλαδή, ακόμη και σε ένα laptop, η ταχύτητα διάγνωσης είναι *τουλάχιστον* γρηγορότερη από τον πραγματικό χρόνο λήψης του βίντεο.



(α) Αγγειοδυσπλασίες



(β) Αιμοφόρα αγγεία



(γ) Φυσιολογικός βλεννογόνος



(δ) Φυσαλίδες

Σχήμα 5.12: Δείγμα εικόνων από το σύνολο δεδομένων



## Κεφάλαιο 6

# Αποτελέσματα

Για την εξαγωγή αποτελεσμάτων, ιδίως σε εφαρμογές νευρωνικών δικτύων, είναι απαραίτητο να δημιουργηθεί ένα καινούργιο σύνολο δεδομένων το οποίο να είναι ξένο με το σύνολο δεδομένων το οποίο χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου. Ο λόγος μπορεί να είναι προφανής, αν αναλογιστεί κανείς ότι ένα νευρωνικό δίκτυο, κατά την εκπαίδευσή του, βρίσκει ένα τοπικό ελάχιστο στο πρόβλημα κατηγοριοποίησης των εικόνων του. Αυτό συμβαίνει πάντα. Το πραγματικό ερώτημα είναι εάν το αποτέλεσμα αυτό μπορεί να γενικευθεί, δηλαδή εάν το τοπικό αυτό ελάχιστο αποτελεί επίσης ένα τοπικό ελάχιστο σε ένα νέο, εντελώς ξένο, σύνολο δεδομένων.

Για τον έλεγχο της μεθόδου, χρησιμοποιήθηκε ένα σύνολο δεδομένων το οποίο περιελάμβανε τρία CE βίντεο ασθενών οι οποίοι διαγνώστηκαν με πολλαπλές αγγειοδυσπλασίες σε όλο το μήκος του λεπτού εντέρου. Από τα βίντεο αυτά απομονώθηκε το κομμάτι κατά το οποίο εικονίζεται το λεπτό έντερο, διάρκειας 2-5 ωρών. Και τα τρία βίντεο ήταν διαστάσεων  $576 \times 576$  ενώ η λήψη τους έγινε με κάψουλες τύπου SB2 και SB3.

Η ευαισθησία (true positive rate - TPR) του αλγορίθμου υπολογίστηκε ως το ποσοστό αγγειοδυσπλασιών που ανιχνεύθηκαν επιτυχώς από τον αλγόριθμο (true positives - TP).

$$TPR = \frac{TP}{P} \cdot 100\% \quad (6.1)$$

όπου  $P$  ο συνολικός αριθμός των αγγειοδυσπλασιών που παρατηρήθηκαν με οπτική εξέταση του βίντεο.

Η ειδικότητα (true negative rate - TNR) υπολογίστηκε ως το ποσοστό των φυσιολογικών εικόνων που ορθώς κατηγοριοποιήθηκαν ως μη-παρεκκλίνουσες (true negatives - TN).

$$TNR = \frac{TN}{N} \cdot 100\% \quad (6.2)$$

όπου, σε αντιστοιχία με πριν,  $N$  είναι ο συνολικός αριθμός των φυσιολογικών εικόνων που εμφανίζονται στο βίντεο.

Για τη μείωση του χρόνου υπολογισμού των τριών βίντεο, σε συνδυασμό με το πρόγραμμα RAPID Viewer το οποίο απορρίπτει διαστήματα κατά τη διάρκεια των οποίων η κάψουλα παρέμεινε στάσιμη, απορρίφθηκαν καρέ τα οποία μπορεί να μην εντοπίστηκαν από το πρόγραμμα παρότι ήταν πανομοιότυπα μεταξύ τους. Για το σκοπό αυτό υπολογίστηκε για κάθε καρέ το ιστογράμμά του και έγινε σύγκριση με το ιστογράμμα του αμέσως προηγούμενου καρέ. Αν τα δύο ιστογράμματα ήταν όμοια σε βαθμό πάνω από κάποιο threshold (99.9%), το καρέ απορριπτόταν και ελεγχόταν το αμέσως επόμενο. Αυτό οδήγησε σε μείωση των εικόνων που ελέγχθηκαν, ο συνολικός αριθμός των οποίων ανήλθε σε 33070.

Κατά την οπτική εξέταση του βίντεο, παρατηρήθηκαν αγγειοδυσπλασίες σε 55, συνολικά, σημεία. Μια αγγειοδυσπλασία θεωρήθηκε ότι ανιχνεύθηκε επιτυχώς εάν ο αλγόριθμος την εντόπισε σε έστω και 1 καρέ μεταξύ όλων όσων εμφανίστηκε. Από τα 55 αυτά σημεία, 51 ανιχνεύθηκαν επιτυχώς από το πρόγραμμα μέσα σε 218 διαφορετικά καρέ. Ευαισθησία, δηλαδή, 92.7%. Άξιο αναφοράς αποτελεί επίσης το γεγονός ότι μερικές από αυτές τις αγγειοδυσπλασίες εμφανίζονταν σε ένα, και μοναδικό, καρέ.

Με κίνδυνο μείωσης του αποτελέσματος της ειδικότητας, θεωρούμε ότι ο πραγματικός αριθμός των μη-φυσιολογικών καρέ είναι ο διπλάσιος από όσα ανιχνεύθηκαν, δηλαδή 436 καρέ. Πρόκειται για μία αρκετά συντηρητική εκτίμηση καθώς οι περισσότερες αγγειοδυσπλασίες ανιχνεύθηκαν επιτυχώς σχεδόν στο σύνολο των καρέ στα οποία εμφανίζονταν, ενώ οι υπόλοιπες ήταν, κυρίως, αγγειοδυσπλασίες οι οποίες εμφανίζονταν για ένα πολύ μικρό χρονικό διάστημα. Με μια οπτική επανεξέταση των βίντεο επιβεβαιώθηκε ότι τα καρέ τα οποία εμφάνιζαν βλάβη και δεν ανιχνεύθηκαν, σε καμία περίπτωση δεν υπερέβαιναν τον αριθμό των ορθώς ανιχνευμένων. Παρόλα αυτά, κρίνεται προτιμότερη η υποτίμηση, παρά η υπερτίμηση του αλγορίθμου. Επομένως, ο αριθμός των φυσιολογικών καρέ ανέρχεται, κατά μία συντηρητική προσέγγιση, στα 32634. Από αυτές τις εικόνες, μόνο 148 ταξινομήθηκαν ως παρεκκλίνουσες. Δηλαδή, η ειδικότητα της μεθόδου σε αυτό το βίντεο ήταν, περίπου, 99.5%.

Τύπος εξέτασης	Αγγειοδυσπ.	Φυσιολ. Βλεν.	TPR (%)	TNR (%)
Με οπτική εξ.	55	32634	—	—
Με χρ. του αλγορ.	51	32486	92.7	99.5

Πίνακας 6.1: Αποτελέσματα για τρία CE βίντεο κάψουλας τεχνολογίας SB2 και SB3 από ασθενείς οι οποίοι διαγνώστηκαν με αγγειοδυσπλασίες.



Σχήμα 6.1: Δείγμα από τις αγγειοδυσπλασίες οι οποίες βρέθηκαν με χρήση του αλγορίθμου στα τρία CE βίντεο.



## Κεφάλαιο 7

# Συμπεράσματα

Στη διπλωματική αυτή εργασία παρουσιάζεται μία νέα μέθοδος αυτόματης ανίχνευσης αγγειοδυσπλασιών. Η μέθοδος αυτή υπολογίζεται σε δύο στάδια. Στο πρώτο στάδιο υπολογίζονται πιθανές περιοχές ενδιαφέροντος οι οποίες κρίνονται πως χρίζουν περαιτέρω διερεύνησης. Στο δεύτερο στάδιο οι περιοχές αυτές ελέγχονται από ένα κατάλληλα εκπαιδευμένο σύστημα μηχανικής μάθησης το οποίο και λαμβάνει την τελική απόφαση για το εάν οι περιοχές αυτές εμφανίζουν βλάβη. Η ευαισθησία αυτής της μεθόδου ανέρχεται σε 92.7%, ενώ η ειδικότητα σε 99.5%.

Όνομα	TPR (%)	TNR (%)
SBI (Yung et. al 2018)	55.7	57.8
Iakovidis et. al (2014)	69.9 – 97.5	–
Iakovidis et. al (2015)	88.0	84.0
Deeba et. al (2018)	94.4	83.9
Παρούσα μελέτη	92.7	99.5

Πίνακας 7.1: Σύγκριση αποτελεσμάτων με άλλες state-of-the-art μελέτες μεθόδων ανίχνευσης.

Για την περαιτέρω βελτίωση της μεθόδου θα πρέπει να εξετασθεί η ευαισθησία του αλγορίθμου επιλογής των περιοχών αυτών. Είναι πιθανό μια εναλλακτική, πιο ευαίσθητη σε αγγειοδυσπλασίες, μέθοδος, να αυξήσει κι άλλο την επίδοση του αλγορίθμου. Εναλλακτικά, θα πρέπει να εξετασθεί η ακρίβεια του νευρωνικού δικτύου βάσει της καταλληλότητας της αρχιτεκτονικής και του συνόλου δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση.

Η υψηλή αυτή ευαισθησία που επιτεύχθηκε, επιβεβαιώνει το γεγονός ότι είναι δυνατός ο περιορισμός της χρήσης των νευρωνικών δικτύων, με σκοπό την εξοικονόμηση υπολογιστικών πόρων, επιλέγοντας μία αρκετά καλή μέθοδο ανίχνευσης περιοχών ενδιαφέροντος. Σε αυτή τη μελέτη, ο αριθμός των περιοχών αυτών περιορίστηκε σε 3, μόνο, ανά εικόνα.

Η ενσωμάτωση της μηχανικής μάθησης στον αλγόριθμο αποτελεί πλεονέκτημα έναντι άλλων μεθόδων, αφού με αυτόν τον τρόπο, για την κατηγοριοποίηση κάθε περιοχής, λαμβάνονται υπ' όψιν, εκτός του χρώματος, και χαρακτηριστικά όπως

είναι το σχήμα, η υφή, ο περιβάλλον βλεπνογόνος, κ.ά. Αυτό επιβεβαιώνεται από την υπερυψηλή ειδικότητα του αλγορίθμου, η οποία ξεπερνά κατά ένα μεγάλο ποσοστό τις υπόλοιπες μεθόδους.

Το γεγονός ότι ο χρόνος εκτέλεσης του αλγορίθμου είναι γρηγορότερος από την καταγραφή των δεδομένων σε πραγματικό χρόνο (334ms ανά εικόνα έναντι 500ms ανά εικόνα) σημαίνει ότι ο αλγόριθμος αυτός μπορεί να ενσωματωθεί στο καταγραφικό μηχάνημα, ώστε η διάγνωση να γίνεται ταυτόχρονα με την καταγραφή των εικόνων.

Βεβαίως, ακόμη και αν κάτι τέτοιο δεν γίνει εφικτό λόγω οικονομικών ή τεχνολογικών περιορισμών, η εφαρμογή της μεθόδου μπορεί να πραγματοποιηθεί με κοινούς ηλεκτρονικούς υπολογιστές, χωρίς την ανάγκη ειδικού ακριβού εξοπλισμού, όπως είναι οι κάρτες γραφικών, αφού ο μέγιστος επιπλέον χρόνος διάγνωσης της κάθε εικόνας υπολογίστηκε σε ένα απλό laptop με κοινά χαρακτηριστικά.

Λόγω της πολύ μεγάλης ακρίβειας της μεθόδου, και του χαμηλού αριθμού ψευδώς θετικών σημείων που προκύπτουν με αυτή, η προτεινόμενη μέθοδος δύναται να συντελέσει αποφασιστικά στη μείωση του χρόνου μελέτης του CE βίντεο για την εξαγωγή διάγνωσης. Μία προανάγνωση του βίντεο με τη χρήση της μεθόδου αυτής η οποία είναι σε θέση να εντοπίσει το 92.7% των αγγειοδυσπλασιών, χωρίς να επιβαρύνει τον ιατρό με επιπλέον άχρηστη πληροφορία λόγω της υψηλής ειδικότητας του 99.5%, μπορεί τόσο να συντομεύσει το χρόνο διάγνωσης, όσο και να ανιχνεύσει αγγειοδυσπλασίες οι οποίες ενδεχομένως να μην παρατηρηθούν από τον ιατρό.

Τέλος, η μέθοδος αυτή ανοίγει νέους ορίζοντες για την εκπαίδευση νευρωνικών δικτύων στην ανίχνευση και άλλων βλαβών του πεπτικού συστήματος, μέσω βελτίωσης των μεθόδων καθορισμού περιοχών ενδιαφέροντος, έτσι ώστε τελικά η διάγνωση των CE βίντεο να πραγματοποιείται αποκλειστικά από έναν ηλεκτρονικό υπολογιστή.

# Βιβλιογραφία

- [1] S. J. Spaner and G. L. Warnock, “A Brief History of Endoscopy, Laparoscopy, and Laparoscopic Surgery”, *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 7, no. 6, pp. 369–373, Dec. 1997.
- [2] G. Levinthal, “The accuracy of an endoscopy nurse in interpreting capsule endoscopy”, *The American Journal of Gastroenterology*, vol. 98, no. 12, pp. 2669–2671, Dec. 2003.
- [3] H. Yamamoto, “Double-Balloon Endoscopy”, *Clinical Gastroenterology and Hepatology*, vol. 3, no. 7, pp. 527–529, 2005.
- [4] M. Delvaux and G. Gay, “Capsule endoscopy: Technique and indications”, *Best Practice & Research. Clinical Gastroenterology*, vol. 22, no. 5, pp. 813–837, 2008.
- [5] C. V. de Bruaene, “Small bowel capsule endoscopy: Where are we after almost 15 years of use?”, *World Journal of Gastrointestinal Endoscopy*, vol. 7, no. 1, p. 13, 2015.
- [6] ASGE Technology Committee, A. Wang, S. Banerjee, B. A. Barth, Y. M. Bhat, S. Chauhan, K. T. Gottlieb, V. Konda, J. T. Maple, F. Murad, P. R. Pfau, D. K. Pleskow, U. D. Siddiqui, J. L. Tokar, and S. A. Rodriguez, “Wireless capsule endoscopy”, *Gastrointestinal Endoscopy*, vol. 78, no. 6, pp. 805–815, Dec. 2013.
- [7] A. Connor, D. Tolan, S. Hughes, N. Carr, and C. Tomson, “Consensus guidelines for the safe prescription and administration of oral bowel-cleansing agents”, *Gut*, vol. 61, no. 11, pp. 1525–1532, Nov. 1, 2012.
- [8] Ε. Χριστοφορίδης, «Ενδοσκοπική Κάψουλα», στο *Ενδοσκοπική Χειρουργική*, σειρά Επεμβατική Ενδοσκόπηση Πεπτικού τόμος Β', Θεσσαλονίκη, 2018, σσ. 253–265.
- [9] A. K. Hara, J. A. Leighton, V. K. Sharma, R. I. Heigh, and D. E. Fleischer, “Imaging of Small Bowel Disease: Comparison of Capsule Endoscopy, Standard Endoscopy, Barium Examination, and CT”, *RadioGraphics*, vol. 25, no. 3, pp. 697–711, May 2005.

- [10] F. Bossa, G. Cocomazzi, M. Valvano, A. Andriulli, and V. Annese, "Detection of abnormal lesions recorded by capsule endoscopy: A prospective study comparing endoscopist's and nurse's accuracy", *Digestive and Liver Disease*, vol. 38, no. 8, pp. 599–602, Aug. 2006.
- [11] T. Omori, T. Hara, S. Sakasai, H. Kambayashi, S. Murasugi, A. Ito, S. Nakamura, and K. Tokushige, "Does the PillCam SB3 capsule endoscopy system improve image reading efficiency irrespective of experience? A pilot study", *Endoscopy International Open*, vol. 6, no. 6, E669–E675, Jun. 2018.
- [12] H. Dokoutsidou, S. Karagiannis, E. Giannakouloupoulou, P. Galanis, N. Kyriakos, C. Liatsos, S. Faiss, and C. Mavrogiannis, "A study comparing an endoscopy nurse and an endoscopy physician in capsule endoscopy interpretation", *European Journal of Gastroenterology & Hepatology*, vol. 23, no. 2, pp. 166–170, Feb. 2011.
- [13] A. Shiotani, K. Honda, M. Kawakami, Y. Kimura, Y. Yamanaka, M. Fujita, H. Matsumoto, K.-i. Tarumi, N. Manabe, and K. Haruma, "Analysis of Small-bowel Capsule Endoscopy Reading by Using Quickview Mode: Training Assistants for Reading May Produce a High Diagnostic Yield and Save Time for Physicians", *Journal of Clinical Gastroenterology*, vol. 46, no. 10, e92–e95, 2012.
- [14] A. Koulaouzidis, A. Smirnidis, S. Douglas, and J. N. Plevris, "QuickView in small-bowel capsule endoscopy is useful in certain clinical settings, but QuickView with Blue Mode is of no additional benefit", *European Journal of Gastroenterology & Hepatology*, vol. 24, no. 9, pp. 1099–1104, Sep. 2012.
- [15] H. Imagawa, S. Oka, S. Tanaka, I. Noda, M. Higashiyama, Y. Sanomura, T. Shishido, S. Yoshida, and K. Chayama, "Improved detectability of small-bowel lesions via capsule endoscopy with computed virtual chromoendoscopy: A pilot study", *Scandinavian Journal of Gastroenterology*, vol. 46, no. 9, pp. 1133–1137, Sep. 2011.
- [16] C. Krystallis, A. Koulaouzidis, S. Douglas, and J. N. Plevris, "Chromoendoscopy in small bowel capsule endoscopy: Blue mode or Fuji Intelligent Colour Enhancement?", *Digestive and Liver Disease: Official Journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver*, vol. 43, no. 12, pp. 953–957, Dec. 2011.
- [17] R. Hain, C. J. Kähler, and C. Tropea, "Comparison of CCD, CMOS and intensified cameras", *Experiments in Fluids*, vol. 42, no. 3, pp. 403–411, Feb. 28, 2007.
- [18] G. Iddan, G. Melon, A. Glukhovskiy, and P. Swain, "Wireless Capsule Endoscopy", *Nature*, vol. 405, pp. 417–418, May 25, 2000.



- [19] I. M. Gralnek, "Obscure-Overt Gastrointestinal Bleeding", *Gastroenterology*, vol. 128, no. 5, pp. 1424–1430, May 2005.
- [20] G. S. Raju, L. Gerson, A. Das, and B. Lewis, "American Gastroenterological Association (AGA) Institute Technical Review on Obscure Gastrointestinal Bleeding", *Gastroenterology*, vol. 133, no. 5, pp. 1697–1717, Nov. 2007.
- [21] S. S. Sami, S. A. Al-Araji, and K. Ragunath, "Review article: Gastrointestinal angiodysplasia - pathogenesis, diagnosis and management", *Alimentary Pharmacology & Therapeutics*, vol. 39, no. 1, pp. 15–34, Jan. 2014.
- [22] M. H. Sleisenger, M. Feldman, L. S. Friedman, and L. J. Brandt, Eds., *Sleisenger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management*, 9th ed, OCLC: ocn318191981, Philadelphia , PA: Saunders/Elsevier, 2010, 2 pp.
- [23] S. J. Boley, R. Sammartano, A. Adams, A. DiBiase, S. Kleinhaus, and S. Sprayregen, "On the nature and etiology of vascular ectasias of the colon. Degenerative lesions of aging", *Gastroenterology*, vol. 72, pp. 650–660, 4 Pt 1 Apr. 1977.
- [24] R. Sidhu, D. S. Sanders, A. J. Morris, and M. E. McAlindon, "Guidelines on small bowel enteroscopy and capsule endoscopy in adults", *Gut*, vol. 57, no. 1, pp. 125–136, Jan. 2008.
- [25] K. T. Tung and A. B. Millar, "Gastric angiodysplasia—a missed cause of gastrointestinal bleeding.", *Postgraduate Medical Journal*, vol. 63, no. 744, pp. 865–866, Oct. 1987.
- [26] P. G. Foutch, D. K. Rex, and D. A. Lieberman, "Prevalence and natural history of colonic angiodysplasia among healthy asymptomatic people", *The American Journal of Gastroenterology*, vol. 90, no. 4, pp. 564–567, Apr. 1995.
- [27] P. H. Duray, J. M. Marcal, V. A. LiVolsi, R. Fisher, C. Scholhamer, and M. H. Brand, "Small intestinal angiodysplasia in the elderly", *Journal of Clinical Gastroenterology*, vol. 6, no. 4, pp. 311–319, Aug. 1984.
- [28] S. Chetcuti Zammit, A. Koulaouzidis, D. S. Sanders, M. E. McAlindon, E. Rondonotti, D. E. Yung, and R. Sidhu, "Overview of small bowel angioectasias: Clinical presentation and treatment options", *Expert Review of Gastroenterology & Hepatology*, vol. 12, no. 2, pp. 125–139, Feb. 2018.
- [29] J. Olmos, M. Marcolongo, V. Pogorelsky, L. Herrera, F. Tobal, and J. R. Davolos, "Push Enteroscopy and Argon Plasma Coagulation in the Prevention of Recurrent Bleeding From Small Bowel Angiodysplasia", *Gastrointestinal Endoscopy*, vol. 61, no. 5, AB177, Apr. 1, 2005.

- [30] N. Takahashi, K. Tanabe, H. Yoshitomi, M. Sato, J. Kitamura, H. Sato, Y. Ishibashi, T. Shimada, and T. Oda, "Successful endoscopic clipping for bleeding from colonic angiodysplasia in a case of Heyde syndrome", *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, vol. 16, no. 9, CS107–109, Sep. 2010.
- [31] T.-Y. Lee, H.-Z. Yeh, S.-S. Yang, and C.-S. Chang, "Successful application of haemoclips plus argon plasma coagulation for angioectasia bleeding with underlying arteriovenous malformation in the colon", *Colorectal Disease: The Official Journal of the Association of Coloproctology of Great Britain and Ireland*, vol. 12, e180–181, 7 Online Jul. 2010.
- [32] J. A. Olmos, M. Marcolongo, V. Pogorelsky, E. Varela, and J. R. Dávalos, "Argon plasma coagulation for prevention of recurrent bleeding from GI angiodysplasias", *Gastrointestinal Endoscopy*, vol. 60, no. 6, pp. 881–886, Dec. 2004.
- [33] T. Marwick and P. Kerlin, "Angiodysplasia of the upper gastrointestinal tract. Clinical spectrum in 41 cases", *Journal of Clinical Gastroenterology*, vol. 8, no. 4, pp. 404–407, Aug. 1986.
- [34] R. R. E. Selinger, G. B. McDonald, D. M. Hockenbery, G. Steinbach, and M. B. Kimmey, "Efficacy of neodymium:YAG laser therapy for gastric antral vascular ectasia (GAVE) following hematopoietic cell transplant", *Bone Marrow Transplantation*, vol. 37, no. 2, pp. 191–197, Jan. 2006.
- [35] B. H. Rogers, "Endoscopic diagnosis and therapy of mucosal vascular abnormalities of the gastrointestinal tract occurring in elderly patients and associated with cardiac, vascular, and pulmonary disease", *Gastrointestinal Endoscopy*, vol. 26, no. 4, pp. 134–138, Nov. 1980.
- [36] A. Bovik, *Handbook of Image and Video Processing*, ser. Academic Press series in communications, networking and multimedia. Austin, Texas: Academic Press, 2000.
- [37] M. Boden, "Clearer Vision", in *Mind as Machine: A History of Cognitive Science*, 2006, p. 781.
- [38] E. Hecht, *Optics*, 4th ed. AW, 2002.
- [39] B. Jähne, *Digital Image Processing*, 6th ed. Springer-Verlag Berlin Heidelberg, 2005.
- [40] L. Shapiro and G. Stockman, *Computer Vision*, 1st ed. Pearson, Feb. 2, 2001, 608 pp.
- [41] D. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, 2nd ed. Boston: Pearson, 2012.
- [42] D. L. MacAdam, "Visual Sensitivities to Color Differences in Daylight\*", *Journal of the Optical Society of America*, vol. 32, no. 5, p. 247, May 1, 1942.

- [43] *Backpropagation*, *Brilliant.org*, Oct. 24, 2018.
- [44] J. Westerhof, J. J. Koornstra, R. A. Hoedemaker, W. J. Sluiter, J. H. Kleibeuker, and R. K. Weersma, “Diagnostic yield of small bowel capsule endoscopy depends on the small bowel transit time”, *World Journal of Gastroenterology*, vol. 18, no. 13, pp. 1502–1507, Apr. 7, 2012.
- [45] S. K. Lo, “How Should We Do Capsule Reading?”, *Techniques in Gastrointestinal Endoscopy*, *Capsule Endoscopy: A Primer for the Endoscopist*, vol. 8, no. 4, pp. 146–148, Oct. 1, 2006.
- [46] M. A. Manfredi, B. K. Abu Dayyeh, Y. M. Bhat, S. S. Chauhan, K. T. Gottlieb, J. H. Hwang, S. Komanduri, V. Konda, S. K. Lo, J. T. Maple, F. M. Murad, U. D. Siddiqui, M. B. Wallace, and S. Banerjee, “Electronic chromoendoscopy”, *Gastrointestinal Endoscopy*, vol. 81, no. 2, pp. 249–261, Feb. 2015.
- [47] S. Han, J. Fahed, and D. R. Cave, “Suspected Blood Indicator to Identify Active Gastrointestinal Bleeding: A Prospective Validation”, *Gastroenterology Research*, vol. 11, no. 2, pp. 106–111, 2018.
- [48] D. E. Yung, C. Sykes, and A. Koulaouzidis, “The validity of suspected blood indicator software in capsule endoscopy: A systematic review and meta-analysis”, *Expert Review of Gastroenterology & Hepatology*, vol. 11, no. 1, pp. 43–51, 2017.
- [49] D. K. Iakovidis and A. Koulaouzidis, “Automatic lesion detection in capsule endoscopy based on color saliency: Closer to an essential adjunct for reviewing software”, *Gastrointestinal Endoscopy*, vol. 80, no. 5, pp. 877–883, Nov. 2014.
- [50] D. K. Iakovidis, D. Chatzis, P. Chrysanthopoulos, and A. Koulaouzidis, “Blood detection in wireless capsule endoscope images based on salient superpixels”, in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan: IEEE, Aug. 2015, pp. 731–734.
- [51] F. Deeba, S. K. Mohammed, F. M. Bui, and K. A. Wahid, “A Saliency-based Unsupervised Method for Angiectasia Detection in Endoscopic Video Frames”, *Journal of Medical and Biological Engineering*, vol. 38, no. 2, pp. 325–335, Apr. 2018.
- [52] D. K. Iakovidis and A. Koulaouzidis, “Software for enhanced video capsule endoscopy: Challenges for essential progress”, *Nature Reviews Gastroenterology & Hepatology*, vol. 12, no. 3, pp. 172–186, Mar. 2015.
- [53] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, “Cluster Analysis: Additional Issues and Algorithms”, in *Introduction to Data Mining*, 2nd ed., Pearson, Jan. 4, 2018, pp. 487–568.

- [54] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-Up Robust Features (SURF)”, *YCVIU Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008, OCLC: 5902435866.
- [55] F.-F. Li, J. Johnson, and S. Yeung, “CS231n: Convolutional Neural Networks for Visual Recognition”, Course Notes, Stanford University, 2018.
- [56] H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki, and S. Carlsson, *From generic to specific deep representations for visual recognition*, 2014.
- [57] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?”, *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.
- [58] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, Dec. 10, 2015.