



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

# Ανίχνευση Φωτιάς σε εικόνα από εναέριο μη επανδρωμένο όχημα UAV με τεχνικές βαθιάς μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΛΥΚΟΓΙΑΝΝΗ ΓΕΩΡΓΙΑΣ**



**Επιβλέπων:** Ανδρέας-Γεώργιος Σταφυλοπάτης  
Καθηγητής

**Συμμετοχή στην Επίβλεψη:** Γεώργιος Αλεξανδρίδης  
Εργαστηριακό και Διδακτικό Προσωπικό

Αθήνα, Ιούνιος 2022

---





# Ανίχνευση Φωτιάς σε εικόνα από εναέριο μη επανδρωμένο όχημα UAV με τεχνικές βαθιάς μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΛΥΚΟΓΙΑΝΝΗ ΓΕΩΡΓΙΑΣ**

**Επιβλέπων:** Ανδρέας-Γεώργιος Σταφυλοπάτης  
Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 9η Ιουνίου 2022.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....  
Ανδρέας-Γεώργιος Σταφυλοπάτης  
Καθηγητής

.....  
Στέφανος Κόλλιας  
Καθηγητής

.....  
Γεώργιος Στάμου  
Καθηγητής





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.

Γεωργία Λυκογιάννη, 2022.

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.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

#### **ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ**

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....

Γεωργία Λυκογιάννη

9η Ιουνίου 2022



## Περίληψη

---

Σκοπός της παρούσας διπλωματικής εργασίας είναι η αντιμετώπιση του φαινομένου της πυρκαγιάς μέσω της έγκαιρης ανίχνευσης της, πριν βγουν εκτός ελέγχου, προκειμένου να ελαχιστοποιηθούν οι επιπτώσεις. Το πρόβλημα των συνεχών δασικών πυρκαγιών, κυρίως κατά τη θερινή περίοδο, αποτελεί ένα σύγχρονο φαινόμενο σε όλο τον κόσμο. Ιδίως, στις χώρες με μεσογειακό κλίμα, όπως η χώρα μας, η κατάσβεση μιας πυρκαγιάς είναι ένα πολύ δύσκολο εγχείρημα.

Έχουν γίνει πολλές προσπάθειες για την επίλυση του συγκεκριμένου προβλήματος. Μερικές από αυτές περιλαμβάνουν τη χρήση αισθητήρων και δορυφορικών εικόνων που λόγω υψηλού κόστους και χρονικής καθυστέρησης δεν ενδείκνυνται ως λύσεις. Μία από τις αναδυόμενες τεχνολογίες για την παρακολούθηση πυρκαγιών αποτελούν τα Μη Επανδρωμένα Εναέρια Οχήματα. Για να υλοποιηθεί ένα τέτοιο σύστημα πυρανίχνευσης απαιτούνται γρήγοροι και υψηλής ακρίβειας αλγόριθμοι επεξεργασίας εικόνας. Με τη ραγδαία πρόοδο της τεχνολογίας, έχει έρθει στο προσκήνιο, η έρευνα με μεθόδους Τεχνητής Νοημοσύνης. Τέτοιου είδους τεχνικές περιλαμβάνουν την εξαγωγή χαρακτηριστικών από εικόνες και βίντεο και τη χρήση βαθύτερων αρχιτεκτονικών, όπως είναι τα σύγχρονα Συνελικτικά Νευρωνικά Δίκτυα.

Αρκετές μελέτες έχουν δείξει ότι η χρήση συνελικτικών τεχνικών είναι στη στάθμη της τεχνικής. Από αυτή την άποψη, προτείνουμε τη χρήση δυο προεκπαιδευμένων αρχιτεκτονικών Xception και MobileNetV2 μάθηση μέσω μεταφοράς. Ακόμη υλοποιήθηκαν και δυο αρχιτεκτονικές κατασκευασμένες από την αρχή που είχαν σαφώς χαμηλότερα αποτελέσματα. Τα αποτελέσματα των δοκιμών δείχνουν ότι η αρχιτεκτονική Xception πετυχαίνει την υψηλότερη ακρίβεια ταξινόμησης των εικόνων στις δύο κλάσεις Πυρκαγιά/Μη-Πυρκαγιά.

## Λέξεις Κλειδιά

Συνελικτικά Νευρωνικά Δίκτυα, Μηχανική Μάθηση, Εναέρια Μη Επανδρωμένα Οχήματα, Ταξινόμηση εικόνων, Αναγνώριση Φωτιάς, Μεταφορά Μάθησης





## Abstract

---

The purpose of this dissertation is to deal with the phenomenon of fire by detecting it on time, before it gets out of control, in order to prevent the consequences. The problem of continuous forest fires prevails all over the world, especially during the summer season. Especially, in countries with a Mediterranean climate, like Greece, fire extinguishing is a very difficult task.

Many efforts have been made to solve this problem. Some of these include the use of sensors and satellite images that are not the best solutions, due to high cost and time delay. One of the emerging technologies for fire monitoring is the use of Unmanned Aerial Vehicles (UAVs). In order for a system to do fire detection, fast and high-precision image processing algorithms are required. With the technology evolution, research is being done with methods of Artificial Intelligence. These techniques include extracting features from images and videos and using deeper architectures, such as Convolutional Neural Networks.

Several studies have shown that the use of convolutional techniques is the current state of the art. From this point of view, we recommend the use of two pre-trained architectures Xception and MobileNetV2 for transfer learning. Moreover, two architectures are built from scratch and they have clearly lower results. The test results show that the Xception architecture achieves the highest image classification accuracy in the two classes Fire / No – Fire.

## Keywords

Convolutional Neural Networks, Machine Learning, Unmanned Aerial Vehicles, Image Classification, Fire Detection, Transfer Learning



*στη μητέρα μου και στην αδερφή μου*



## Ευχαριστίες

---

Με την ολοκλήρωση της διπλωματικής μου εργασίας, περατώνονται και οι προπτυχιακές σπουδές μου στο τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Κατά τη διάρκεια αυτών των χρόνων, εκτός από τη δική μου προσπάθεια, συντέλεσαν καθοριστικά και πρόσωπα που ήταν δίπλα μου και με εμπύχωναν συνεχώς.

Καταρχάς θα ήθελα να ευχαριστήσω τον επιβλέποντα κ. Ανδρέα-Γεώργιο Σταφυλοπάτη, Καθηγητή Ε.Μ.Π., για την εμπιστοσύνη και την ευκαιρία που μου έδωσε να εκπονήσω τη διπλωματική μου στο Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης. Επίσης, θα ήθελα να ευχαριστήσω τα μέλη της επιτροπής, κ.κ. Κόλλια Στέφανο και Στάμου Γεώργιο, Καθηγητές Ε.Μ.Π., για την τιμή που μου έκαναν να συμμετάσχουν σε αυτή. Τέλος, θα ήθελα να ευχαριστήσω ιδιαίτερα τον συνεπιβλέποντα κ. Γεώργιο Αλεξανδρίδη, Ε.ΔΙ.Π. Ε.Μ.Π, για την καθοδήγησή του και την εξαιρετική συνεργασία που είχαμε.

Ταυτόχρονα, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου. Η στήριξη τους και η βοήθεια που μου προσέφεραν όλα αυτά τα χρόνια ήταν πολύτιμη και διαρκής.

Αθήνα, Ιούνιος 2022

*Γεωργία Λυκογιάννη*



# Περιεχόμενα

---

<b>Περίληψη</b>	<b>1</b>
<b>Abstract</b>	<b>3</b>
<b>Ευχαριστίες</b>	<b>7</b>
<b>1 Εισαγωγή</b>	<b>15</b>
1.1 Το φαινόμενο της φωτιάς . . . . .	15
1.2 Φωτιές και παγκόσμια κλιματική αλλαγή . . . . .	16
1.3 Κίνητρο και αντικείμενο της διπλωματικής . . . . .	17
1.4 Δομή της Εργασίας . . . . .	18
<b>2 Θεωρητικό υπόβαθρο</b>	<b>19</b>
2.1 Μηχανική Μάθηση . . . . .	19
2.1.1 Επιβλεπόμενη μάθηση (Supervised Learning) . . . . .	19
2.1.2 Μη επιβλεπόμενη μάθηση (Unsupervised Learning) . . . . .	20
2.1.3 Ενισχυτική Μάθηση (Reinforcement Learning) . . . . .	20
2.2 Τεχνητά Νευρωνικά Δίκτυα . . . . .	21
2.3 Βαθιά Νευρωνικά Δίκτυα . . . . .	23
2.4 Το πρόβλημα της υπερπροσαρμογής και της υποπροσαρμογής . . . . .	23
2.5 Συνελκτικά Νευρωνικά Δίκτυα (CNNs) . . . . .	26
2.5.1 Ορισμός . . . . .	26
2.5.2 Επίπεδα Συνελκτικών Νευρωνικών Δικτύων . . . . .	27
2.5.3 Παράμετροι και Διαστάσεις Εξόδων . . . . .	29
2.5.4 Η Εξέλιξη των Συνελκτικών Νευρωνικών Δικτύων . . . . .	31
<b>3 Σχετικές εργασίες και περιγραφή επιλεγμένων αρχιτεκτονικών</b>	<b>35</b>
3.1 Σχετικές εργασίες . . . . .	35
3.1.1 Κλασικές προσεγγίσεις . . . . .	35
3.1.2 Σύγχρονες προσεγγίσεις . . . . .	38
3.1.3 Προσεγγίσεις βαθιάς μάθησης . . . . .	39
3.2 Μεταφορά Μάθησης . . . . .	41
3.2.1 Xception Model . . . . .	42
3.2.2 MobileNetV2 Model . . . . .	45
3.3 Βελτιστοποίηση Υπερπαραμέτρων . . . . .	49
3.3.1 Αλγόριθμοι Βελτιστοποίησης και Ρυθμός Μάθησης . . . . .	50

3.3.2 Μέγεθος δέσμης . . . . .	52
3.3.3 Εποχές . . . . .	52
3.3.4 Πλήθος των Πλήρως διασυνδεδεμένων επιπέδων . . . . .	52
3.3.5 Αρχικοποιητές πυρήνα . . . . .	52
3.3.6 Ποσοστό εγκατάλειψης . . . . .	52
<b>4 Σύνολο Δεδομένων και Εργαλεία</b>	<b>53</b>
4.1 Δεδομένα . . . . .	53
4.2 Εργαλεία . . . . .	54
<b>5 Πρακτική υλοποίηση και αποτελέσματα</b>	<b>57</b>
5.1 Αξιολόγηση Απόδοσης . . . . .	57
5.2 Υλοποιήσεις με μεταφορά μαθησης . . . . .	58
5.3 Υλοποιήσεις με κατασκευή απο την αρχή . . . . .	63
5.4 Συγκεντρωτικά Αποτελέσματα . . . . .	67
<b>6 Συμπεράσματα και Μελλοντικές Κατευθύνσεις</b>	<b>69</b>
6.1 Συμπεράσματα . . . . .	69
6.2 Μελλοντικές Επεκτάσεις . . . . .	70
<b>Βιβλιογραφία</b>	<b>79</b>
<b>Απόδοση ξενόγλωσσων όρων</b>	<b>81</b>



## Κατάλογος Εικόνων

---

2.1	Νευρώνας Ανθρώπινου Εγκεφάλου (Πηγή [1]) . . . . .	21
2.2	Δομή Τεχνητού Νευρωνικού Δικτύου (Πηγή [2]) . . . . .	22
2.3	Διάγραμμα που αναπαριστά τη σχέση μεταξύ της τεχνητής νοημοσύνης, της μηχανικής μάθησης, των νευρωνικών δικτύων, της βαθιάς μάθησης και των συνελκτικών νευρωνικών δικτύων. Η τεχνητή νοημοσύνη είναι η πιο γενική έννοια καθώς περιλαμβάνει συστήματα που αποσκοπούν να μιμηθούν την ανθρώπινη ευφυΐα. . . . .	23
2.4	Αναπαράσταση των προβλημάτων υπερπροσαρμογής και υποπροσαρμογής (Πηγή [3]) . . . . .	24
2.5	Τεχνικές Data Augmentation (Πηγή [4]) . . . . .	25
2.6	Τεχνική Dropout (Πηγή [5]) . . . . .	25
2.7	Παράδειγμα συνελκτικού νευρωνικού δικτύου (Πηγή [6]) . . . . .	27
2.8	Εφαρμογή συνέλιξης φίλτρου σε εικόνα εισόδου (Πηγή [7]) . . . . .	28
2.9	Παράδειγμα υποδειγματοληψίας μεγίστου και μέσου όρου (Πηγή [8]) . . . . .	28
2.10	Παράδειγμα μετατροπής σε μονοδιάστατη έξοδο (Πηγή [9]) . . . . .	29
2.11	Αρχιτεκτονική LeNet (Πηγή [10]) . . . . .	31
2.12	Αρχιτεκτονική AlexNet (Πηγή [11]) . . . . .	32
2.13	Αρχιτεκτονική GoogleNet (Πηγή [12]) . . . . .	33
2.14	Inception μονάδα της GoogleNet (Πηγή [12]) . . . . .	33
2.15	Αρχιτεκτονική VGG16 (Πηγή [13]) . . . . .	34
2.16	Αντιπαράθεση ενός απλού δικτύου 34 επιπέδων και ενός residual δικτύου (Πηγή [14]) . . . . .	34
3.1	Αντιπαράθεση της μάθησης μέσω μεταφοράς και της εκπαίδευσης από την αρχή (Πηγή [15]) . . . . .	42
3.2	Παράδειγμα της κατά βάθος διαχωρίσιμης συνέλιξης, όπου ο χάρτης εισόδου έχει διαστάσεις $8 \times 8 \times 3$ , ο πυρήνας έχει διαστάσεις $3 \times 3 \times 3$ και ο χάρτης εξόδου $8 \times 8 \times 1$ (Πηγή [16]). . . . .	43
3.3	Αρχιτεκτονική Xception δικτύου (Πηγή [17]) . . . . .	45
3.4	Δομικά στοιχεία αρχιτεκτονικής MobileNetV1 (Πηγή [18]) . . . . .	46
3.5	Απεικόνιση μπλοκ συμφόρησης με κατάλοιπο. Παράδειγμα block με $n > 1$ και $s = 2$ . Στα δεξιά απεικονίζεται η δομή του πρώτου block και στα αριστερά η δομή όλων των μετέπειτα blocks (Πηγή [19]). . . . .	46
3.6	Αντιπαράθεση του υπολειπόμενου μπλοκ και του ανεστραμμένου υπολειπόμενου μπλοκ (Πηγή [20]) . . . . .	48

3.7	Αρχιτεκτονική του MobileNetV2 (Πηγή [21]) . . . . .	49
4.1	Παραδείγματα εικόνων φωτιάς από το training set (Πηγή [22]) . . . . .	54
4.2	Παραδείγματα εικόνων χωρίς φωτιά από το test set (Πηγή [22]) . . . . .	55
5.1	Πίνακας σύγκρισης δύο κλάσεων (Πηγή [23]) . . . . .	57
5.2	Πίνακας σύγκρισης πρώτου μοντέλου . . . . .	60
5.3	Διάγραμμα του πρώτου μοντέλου για την πιστότητα και την απώλεια στα σύνολα εκπαίδευσης και επικύρωσης ως προς τις εποχές . . . . .	60
5.4	Πίνακας σύγκρισης δεύτερου μοντέλου . . . . .	62
5.5	Διάγραμμα του τρίτου μοντέλου για την πιστότητα και την απώλεια στα σύνολα εκπαίδευσης και επικύρωσης ως προς τις εποχές . . . . .	62
5.6	Τρίτο μοντέλο . . . . .	63
5.7	Διάγραμμα του τρίτου μοντέλου για την πιστότητα και την απώλεια στα σύνολα εκπαίδευσης και επικύρωσης ως προς τις εποχές . . . . .	64
5.8	Πίνακας Σύγκρισης του τρίτου μοντέλου . . . . .	64
5.9	Τέταρτο μοντέλο . . . . .	65
5.10	Διάγραμμα του τέταρτου μοντέλου για την πιστότητα και την απώλεια στα σύνολα εκπαίδευσης και επικύρωσης ως προς τις εποχές . . . . .	66
5.11	Πίνακας Σύγκρισης του τέταρτου μοντέλου . . . . .	67

## Κατάλογος Πινάκων

---

3.1	Περιγραφή μπλοκ συμφόρησης με κατάλοιπο (Πηγή [21]) . . . . .	48
5.1	Πίνακας υπερπαραμέτρων για τη μεταφορά μάθησης με πρώτο μοντέλο Keras – <b>Xception</b> . . . . .	59
5.2	Πίνακας υπερπαραμέτρων για τη μάθηση μέσω μεταφοράς με το δεύτερο μο- ντέλο Keras – <b>MobileNetV2</b> . . . . .	61
5.3	Μετρικές αξιολόγησης και παράμετροι όλων των μοντέλων . . . . .	68



## Κεφάλαιο **1**

### Εισαγωγή

---

**Ο**ι πυρκαγιές είναι μια από τις πιο θανατηφόρες φυσικές καταστροφές σε όλη την υφήλιο. Προκαλούν ζημιές σε εκατομμύρια στρέμματα δασικών εκτάσεων και οικουσθήματα, απειλούν τις ζωές άγριων ζώων και θέτουν σε κίνδυνο ανθρώπινες ζωές και περιουσίες, ειδικά όταν οι φωτιές εξελίσσονται σε κατοικημένες περιοχές. Σημαντικός είναι και ο κίνδυνος που διατρέχουν οι πυροσβέστες και οι επιχειρησιακές δυνάμεις. Ως κοινωνία θέλουμε και απαιτούμε να δημιουργηθεί ένα έξυπνο σύστημα το οποίο θα ανιχνεύει έγκαιρα τις πυρκαγιές πριν βγουν εκτός ελέγχου.

#### 1.1 Το φαινόμενο της φωτιάς

Το φαινόμενο της φωτιάς ορίζεται ως η καύση η οποία συνοδεύεται από φλόγα. Είναι η χημική αντίδραση κατά την οποία εκλύεται έντονη θερμότητα και η φλόγα είναι το ορατό αποτέλεσμα της εξώθερμης αντίδρασης. Η φωτιά είναι θερμή λόγω της μετατροπής του αδύναμου διπλού δεσμού σε μοριακό οξυγόνο,  $O_2$ , στους ισχυρότερους δεσμούς στα προϊόντα καύσης διοξείδιο του άνθρακα και το νερό απελευθερώνει ενέργεια (418 kJ ανά 32 g  $O_2$ ). Οι ενεργειακές σχέσεις του καυσίμου παίζουν πολύ μικρό ρόλο. Οι φλόγες παράγονται στο σημείο ανάφλεξης και αποτελούνται κυρίως από διοξείδιο του άνθρακα, υδρατμούς, οξυγόνο και άζωτο. Εφόσον τα αέρια είναι αρκετά ζεστά, μπορεί να ιονιστούν για την παραγωγή πλάσματος. Ανάλογα με την ουσία που ανάβει, και τυχόν άλλα στοιχεία που υπάρχουν κοντά σε αυτήν, το χρώμα της φλόγας και η ένταση της φωτιάς μπορεί να είναι διαφορετικά.

Η φωτιά συνήθως οδηγεί σε πυρκαγιά και έχει τη δυνατότητα να προκαλέσει φυσική καταστροφή. Η πυρκαγιά, είναι ένας από τους σημαντικότερους κινδύνους των δασών, και τις περισσότερες φορές είναι αποτέλεσμα ανθρώπινης δραστηριότητας είτε από αμέλεια ή εμπρησμό. Όπως αναφέρθηκε και προηγουμένως, μπορεί να καταστρέψει στρέμματα καλυμμένα με βλάστηση, φυσικά ενδιαίτηματα ζώων και πουλιών, να απελευθερώσει ρύπους στην ατμόσφαιρα, να διαταράξει τον υδρολογικό κύκλο μιας περιοχής, να απειλήσει κατοικίες, και να θέσει σε κίνδυνο ανθρώπινες ζωές. Μετά από την καταστροφή των δασών και της βλάστησης, επόμενο είναι οι έντονες βροχοπτώσεις που μπορούν να οδηγήσουν σε διάβρωση του εδάφους από το νερό.

## 1.2 Φωτιές και παγκόσμια κλιματική αλλαγή

Η έναρξη και η εξάπλωση μιας πυρκαγιάς εξαρτάται κατά πολύ από τους κλιματικούς παράγοντες δηλαδή τη θερμοκρασία, την υγρασία, την ταχύτητα του ανέμου που επικρατούν στην περιοχή, τόσο κατά τη στιγμή έναρξης, όσο και γενικότερα. Οι αλλαγές που έχουν υποστεί οι κλιματικοί παράγοντες, γνωστές ως κλιματική αλλαγή, είναι ακόμα μια συνέπεια των πυρκαγιών, οι οποίοι με τη σειρά τους, επηρεάζουν την εμφάνιση των πυρκαγιών, δημιουργώντας μια αμφίδρομη σχέση μεταξύ της κλιματικής αλλαγής και της φωτιάς.

Με βάση τις αναφορές του Εθνικού Διυπηρεσιακού Πυροσβεστικού Κέντρου (NIFC) στις ΗΠΑ, ο συνολικός αριθμός των πυρκαγιών στα δάση ήταν, κατά μέσο όρο ετησίως, 62.769 (από το 2011 έως το 2020) κατά τη διάρκεια των οποίων κάηκαν περισσότερα από 6.505.597 στρέμματα γης (ετησίως κατά μέσο όρο), αντιστοιχώντας σε ζημιές άνω των 6 δισεκατομμυρίων δολαρίων [24, 25].

Οι εκπομπές διοξειδίου του άνθρακα κινήθηκαν σε επίπεδα ρεκόρ το 2021 από τις τεράστιες φωτιές που ξέσπασαν. Σύμφωνα με τον Mark Parrington, ανώτερο επιστήμονα της Υπηρεσίας Παρακολούθησης της Ατμόσφαιρας “Copernicus”, οι πυρκαγιές έως το περασμένο καλοκαίρι, του 2021, απελευθέρωσαν περίπου 4,7 γιγατόνους διοξειδίου του άνθρακα. Τα στοιχεία προβληματίζουν τους επιστήμονες, λόγω των επιπτώσεων του CO<sub>2</sub> που συμβάλλει στην κλιματική αλλαγή.

Το φετινό ρεκόρ στην εκπομπή διοξειδίου του άνθρακα οφείλεται ως επί το πλείστον στις περισσότερες από 200 δασικές πυρκαγιές που κατέστρεψαν τμήματα της Σιβηρίας. Σύμφωνα με τη NASA, μόνο τον Αύγουστο του 2021, εκπέμφθηκαν 505 μεγατόνοι διοξειδίου του άνθρακα πάνω από τη βορειοανατολική Σιβηρία μέχρι την απομακρυσμένη Δημοκρατία της Σάχα. Οι τεράστιοι μαύροι, πυκνοί καπνοί από εκατοντάδες δασικές πυρκαγιές διασκορπίστηκαν 2.000 μίλια από ανατολικά προς δυτικά και 2.000 μίλια από βορρά προς νότο. Ο καπνός έφτασε στον Βόρειο Πόλο και εξαπλώθηκε σε μέρη της Μογγολίας, σε απόσταση 1.200 μιλίων. Ο Parrington ανέφερε ότι μόνο τον Ιούλιο μήνα εκλύθηκαν 350 μεγατόνοι διοξειδίου του άνθρακα, επίπεδα ρεκόρ μήνα από τότε που οι δορυφόροι άρχισαν να καταγράφουν τα επίπεδα CO<sub>2</sub>. Επίσης, επισήμανε ότι οι εκπομπές τροφοδοτήθηκαν από τον καύσωνα και τις παρατεταμένες ξηρασίες, που οφείλονται στην κλιματική αλλαγή.

Εκτός από τις Ηνωμένες Πολιτείες της Αμερικής και τη Ρωσία, τον περασμένο χρόνο εκδηλώθηκαν τεράστιες πυρκαγιές στη Νότια Ευρώπη. Η Ελλάδα, η Ιταλία, η Τουρκία, η Ισπανία, η Πορτογαλία και το Μαυροβούνιο έχουν υποστεί μεγάλες ζημιές, όπως και στη Μέση Ανατολή: η Αλγερία, ο Λίβανος και η Τυνησία.

Το πρόβλημα παρέμεινε έντονο και στο δάσος του Αμαζονίου, όπου μέχρι τον Αύγουστο εκδηλώθηκαν 267 μεγάλες φωτιές καίγοντας 150.000 στρέμματα, μια περιοχή στο μέγεθος του Λος Άντζελες. Στις 27 Ιουνίου, η κυβέρνηση της Βραζιλίας επέβαλε απαγόρευση στις μη εγκεκριμένες φωτιές σε εξωτερικούς χώρους, αλλά οι παρελθούσες πολιτικές είχαν ήδη συμβάλει στην αύξηση των φωτιών στην Αμαζονία μεταξύ 2018-2019, σύμφωνα με το Εθνικό Ινστιτούτο για την Έρευνα του Διαστήματος της Βραζιλίας.

Η Ελλάδα είναι μια χώρα με υψηλό βαθμό επικινδυνότητας πυρκαγιάς, καθώς το μεσογειακό κλίμα θεωρείται επιρρεπές στις φωτιές. Οι τόποι που συνδέονται με το μεσογειακό κλίμα, είναι περιοχές στις οποίες η βροχόπτωση υπερβαίνει τη δυνητική εξατμισοδιαπνοή

κατά τη διάρκεια της εποχής των βροχών, με αποτέλεσμα την επαρκή ανάπτυξη φυτών, που γίνονται ιδιαίτερα εύφλεκτα κατά την διάρκεια της ξηρής περιόδου του καλοκαιριού. Στην Ελλάδα τα ενδημικά δασικά είδη είναι πυρόφιλα είδη, με χαρακτηριστικό παράδειγμα τέτοιων φυτών τα πευκα.

Το 2021 είναι η χρονιά που οι επιπτώσεις της κλιματικής αλλαγής, έκαναν ακόμη πιο αισθητή την παρουσία τους και στη χώρα μας. Είναι η χρονιά που βιώσαμε επικίνδυνα φαινόμενα, είναι η χρονιά που κήκε η Ελλάδα, με το νησί της Εύβοιας να βρίσκεται στο επίκεντρο. Ανάμεσα στις 428 δασικές πυρκαγιές που κατέγραψε η Πυροσβεστική Υπηρεσία το περασμένο καλοκαίρι, εκείνη της βόρειας Εύβοιας ήταν η πλέον καταστροφική, σε μια περιοχή με μοναδικό φυσικό πλούτο. Οι πυρκαγιές επιδεινώθηκαν λόγω των πολύ υψηλών, μη φυσιολογικών θερμοκρασιών. Ως επακόλουθα αυτών, πλημμύρες και ακραία φαινόμενα έπληξαν τη χώρα, αφήνοντας πίσω τους νεκρούς και τεράστιες καταστροφές. Αξίζει να αναφερθεί μια πολύ καταστροφική πυρκαγιά των τελευταίων ετών, που ήταν στο Μάτι Αττικής τον Ιούλιο του 2018, η οποία εκτός των άλλων στοίχισε και τη ζωή σε 102 ανθρώπους.

Όλα αυτά τα ανησυχητικά γεγονότα παρακινούν τους ερευνητές να αναζητήσουν νέες λύσεις για την έγκαιρη ανίχνευση και διαχείριση πυρκαγιάς.

### 1.3 Κίνητρο και αντικείμενο της διπλωματικής

Οι πρόσφατες εξελίξεις στα συστήματα εναέριας παρακολούθησης μπορούν να παρέχουν στους πρώτους ανταποκριτές και στις επιχειρησιακές δυνάμεις πιο ακριβή δεδομένα σχετικά με τη συμπεριφορά της πυρκαγιάς για βελτιωμένη διαχείριση της πυρκαγιάς.

Οι πυρκαγιές εκδηλώνονται συχνά σε απομακρυσμένες δασικές περιοχές όπου οι κοινές μέθοδοι ανίχνευσης, όπως οι σταθμοί πύργων επιφυλακής, αποτυγχάνουν να εντοπίσουν τέτοιες πυρκαγιές εγκαίρως. Επίσης, οι συμβατικές προσεγγίσεις ανίχνευσης μετά βίας μπορούν να παρέχουν επαρκείς πληροφορίες σχετικά με τις ακριβείς θέσεις της και τον προσανατολισμό της επέκτασης της πυρκαγιάς. Πρόσφατες έρευνες προτείνουν καινοτομίες (λ.χ. λύσεις Internet of Things - IoT) που βασίζονται σε ασύρματα δίκτυα αισθητήρων [26, 27, 28], αλλά τέτοια δίκτυα έχουν υψηλό κόστος και χρειάζονται πολλές δοκιμές πριν από την παροχή πρακτικών πληροφοριών. Η ανίχνευση πυρκαγιάς με χρήση ασύρματων δικτύων αισθητήρων, εκτός από δαπανηρή, έχει και υψηλή συντήρηση για την κάλυψη μεγάλων δασικών εκτάσεων [29]. Οι αισθητήρες είναι είτε υπέρυθρης ακτινοβολίας ή αισθητήρες καπνού οι οποίοι ανιχνεύουν την παρουσία ορισμένων σωματιδίων που παράγονται από τον ιονισμό φωτιάς και όχι εξαιτίας της ίδιας της καύσης. Συνεπώς, παρουσιάζουν συχνά χρονική καθυστέρηση και έχουν λανθασμένα αποτελέσματα. Σε ευρύτερες κλίμακες χρησιμοποιούνται οι δορυφορικές εικόνες για την αξιολόγηση των πυρκαγιών [30]. Ωστόσο, οι δορυφόροι δεν μπορούν να παρέχουν βίντεο ή εικόνες σε πραγματικό χρόνο, αφού η ποιότητα των εικόνων τους επηρεάζεται σε μεγάλο βαθμό από τις καιρικές συνθήκες. Τα επανδρωμένα αεροσκάφη μπορούν να ερευνηθούν με ακρίβεια μια ευρεία περιοχή σε σύντομο χρονικό διάστημα, ωστόσο, αυτή η λύση είναι επίσης δαπανηρή και μπορεί να θέσει σε κίνδυνο τη ζωή των πιλότων λόγω της ροής αέρα υψηλής θερμοκρασίας και του πυκνού καπνού.

Λόγω των συνεχών προκλήσεων και τα ζητήματα των προηγούμενων μεθόδων, η χρήση μη επανδρωμένων εναέριων οχημάτων (UAV) με ενσωματωμένη κάμερα, για την παρακολο-

ύθηση πυρκαγιάς, κερδίζει όλο και μεγαλύτερο έδαφος τα τελευταία χρόνια. Τα UAV, ως μια χαμηλού κόστους και ευέλικτη λύση για τη συλλογή δεδομένων, λαμβάνουν υπόψη και τα μοναδικά χαρακτηριστικά τους όπως τρισδιάστατες κινήσεις, εύκολη πτήση και ευελιξία [31, 32, 33, 34, 35]. Πρόσφατες μελέτες διερεύνησαν τη χρήση UAV σε επιχειρήσεις φυσικών καταστροφών όπως πυρκαγιές και πλημμύρες, ιδιαίτερα ως προσωρινή λύση όταν τα επίγεια δίκτυα αποτυγχάνουν λόγω κατεστραμμένων υποδομών και προβλημάτων επικοινωνίας [36, 37, 38]. Μπορούν να προσφέρουν δυνατότητες όπως παρακολούθηση της γραμμής του μετώπου πυρκαγιάς, γρήγορη χαρτογράφηση περιοχών και εκτίμηση ζημιών, ροή βίντεο σε πραγματικό χρόνο και αναζήτηση για διάσωση ζώων [39, 40, 41, 42].

Ο χρόνος απόκρισης σε ένα τέτοιο σύστημα μπορεί να είναι γρηγορότερος από τους παραδοσιακούς αισθητήρες, καθώς οι κάμερες δεν χρειάζεται να περιμένουν να διαχυθεί ο καπνός ή η θερμότητα. Έτσι, θα ενεργοποιείται άμεσα η επέμβαση της πυροσβεστικής για την κατάσβεσή της όταν είναι ακόμα σε πρώιμο στάδιο ώστε να αποφευχθεί η εξάπλωση της.

Σε αυτή τη διπλωματική, εστιάζουμε στην έγκαιρη ανίχνευση πυρκαγιών από συλλεγμένες έγχρωμες εικόνες χρησιμοποιώντας UAV για την ανάπτυξη ενός μηχανισμού χαμηλού υπολογισμού κατάλληλου για μη επανδρωμένα εναέρια οχήματα με περιορισμένους πόρους. Ως εκ τούτου, αναπτύσσουμε έναν μηχανισμό ανίχνευσης πυρκαγιάς που βασίζεται σε συνελκτικά νευρωνικά δίκτυα.

## 1.4 Δομή της Εργασίας

Η εργασία αυτή είναι οργανωμένη σε έξι κεφάλαια. Στο παρόν κεφάλαιο έγινε μια εισαγωγή στο πρόβλημα των πυρκαγιών και στο τι έπεται να μελετηθεί. Στο επόμενο κεφάλαιο περιγράφεται το προαπαιτούμενο θεωρητικό υπόβαθρο για την κατανόηση των προτεινόμενων δικτύων και δίνεται βάση στα Συνελκτικά Νευρωνικά Δίκτυα. Στο Κεφάλαιο 3, περιγράφονται οι σχετικές εργασίες της βιβλιογραφίας με το θέμα και στη συνέχεια αναλύονται οι αρχιτεκτονικές που χρησιμοποιήθηκαν για την επίλυση του προβλήματος. Στο Κεφάλαιο 4 παρατίθενται η βάση δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση των μοντέλων και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στο Κεφάλαιο 5 παρουσιάζεται η πρακτική υλοποίηση και τα πειραματικά αποτελέσματα. Τέλος, στο Κεφάλαιο 6 δίνονται τα συμπεράσματα της διπλωματικής εργασίας, καθώς και μελλοντικές επεκτάσεις.



# Θεωρητικό υπόβαθρο

---

## 2.1 Μηχανική Μάθηση

Η μηχανική μάθηση αποτελεί κλάδο της τεχνητής νοημοσύνης. Το 1959, ο Arthur Samuel εισάγει τον όρο ‘μηχανική μάθηση’ και την ορίζει ως ‘το ερευνητικό πεδίο που δίνει τη δυνατότητα στους υπολογιστές να μαθαίνουν χωρίς να έχουν προγραμματιστεί ρητά’ [43]. Ένας ορισμός για τη μηχανική μάθηση δόθηκε και από τον Tom M. Mitchell το 1997: ‘Η μηχανική μάθηση είναι η μελέτη των αλγορίθμων που επιτρέπουν στα προγράμματα υπολογιστών να βελτιώνονται αυτόματα μέσω της εμπειρίας’ [44].

Είναι ένας ευρύς όρος που αναφέρεται σε υπολογιστικούς αλγόριθμους που βασίζονται σε εμπειρία για τη βελτίωση της απόδοσης και τη δημιουργία ακριβών προβλέψεων. Στην περίπτωση αυτή, η εμπειρία αναφέρεται στην προηγούμενη γνώση του εκπαιδευτή, η οποία είναι συχνά σε μορφή ηλεκτρονικών δεδομένων που συλλέγονται και διατίθενται για ανάλυση. Αυτές οι πληροφορίες θα μπορούσαν να συλλέγονται μέσω της αλληλεπίδρασης με το περιβάλλον. Σε κάθε περίπτωση, η ποιότητα και το μέγεθος είναι κρίσιμα για την επιτυχία των προβλέψεων. Προφανώς, όσο μεγαλύτερο είναι το μέγεθος του δείγματος, τόσο πιο εύκολη είναι η εργασία.

Η μηχανική μάθηση είναι άρρηκτα συνδεδεμένη με την ανάλυση δεδομένων και τη στατιστική, καθώς η αποτελεσματικότητα ενός αλγορίθμου μάθησης καθορίζεται από τα δεδομένα που χρησιμοποιούνται. Οι τεχνικές μάθησης, γενικά, είναι μεθοδολογίες που βασίζονται σε δεδομένα και συνδυάζουν θεμελιώδεις έννοιες της επιστήμης των υπολογιστών με στατιστικές έννοιες. Μπορεί να χρησιμοποιηθεί σε ένα ευρύ φάσμα εφαρμογών και μερικά παραδείγματα αυτών είναι: η επεξεργασία φυσικής γλώσσας, η οπτική αναγνώριση χαρακτήρων, η όραση υπολογιστών [45].

Χωρίζεται σε τρεις κατηγορίες ανάλογα με τη φύση και τον στόχο του εκάστοτε προβλήματος. Ας σημειώσουμε, ότι οι κατηγορίες αυτές είναι, σε μεγάλο βαθμό, συνυφασμένες με τους τρόπους που μαθαίνει ο άνθρωπος. Οι κατηγορίες θα αναλυθούν παρακάτω.

### 2.1.1 Επιβλεπόμενη μάθηση (Supervised Learning)

Αναφέρεται επίσης ως μάθηση με εκπαιδευτή. Μπορεί να θεωρηθεί ότι ο εκπαιδευτής έχει γνώση του περιβάλλοντος και αυτή αντιπροσωπεύεται από ένα σύνολο παραδειγμάτων εισόδου-εξόδου. Οι αλγόριθμοι κατασκευάζουν μια συνάρτηση απεικόνισης των δεδομένων εισόδου με τις γνωστές εξόδους τους, ώστε να γίνει μια επιτυχημένη γενίκευση της συνάρ-

τησης αυτής και να προβλέπονται για νέα δεδομένα εισόδου οι επιθυμητές και άγνωστες εξόδους τους. Συνεπώς, για την εκπαίδευση απαιτούνται δεδομένα με επισημασμένες τις σωστές απαντήσεις τους. Αυτή η μάθηση συναντάται κυρίως σε προβλήματα ταξινόμησης, πρόβλεψης και διερμηνείας.

Στην επιβλεπόμενη μάθηση έχουμε δύο υποκατηγορίες προβλημάτων ανάλογα με την τιμή της ετικέτας, δηλαδή αν είναι ποιοτική ή ποσοτική. Πιο συγκεκριμένα, έχουμε τις παρακάτω:

- Ταξινόμηση (Classification): Όταν τα δεδομένα εισόδου χωρίζονται σε δύο ή περισσότερες κλάσεις, στις οποίες το μοντέλο που εκπαιδεύεται πρέπει να τα αντιστοιχίσει σωστά. Ως έξοδος προβλέπεται μία διακριτή τιμή (ποιοτική ετικέτα) ως ταμπέλα μιας από τις κλάσεις του προβλήματος.
- Παλινδρόμηση (Regression): Όταν το εκπαιδευμένο μοντέλο προβλέπει στην έξοδο του συνεχείς αριθμητικές τιμές (ποσοτική τιμή ετικέτας). Τα μοντέλα αυτά στοχεύουν στη πρόβλεψη μίας τιμής ενός μεγέθους για το οποίο έχουν στατιστικά ή ιστορικά δεδομένα. Συναντώνται συχνά στον τομέα των πωλήσεων, καθώς μπορούν να προβλέψουν ποσοστά πωλήσεων ενός προϊόντος.

### 2.1.2 Μη επιβλεπόμενη μάθηση (Unsupervised Learning)

Αναφέρεται και ως αυτό-οργανούμενη μάθηση καθώς δεν υπάρχει εξωτερικός εκπαιδευτής που να επιβλέπει τη μάθηση. Οι αλγόριθμοι, βάσει των δεδομένων εισόδου, κατασκευάζουν μοντέλα μέσω των οποίων εντοπίζονται συσχετίσεις και δημιουργούνται ομάδες δεδομένων με σκοπό την επίλυση του προβλήματος χωρίς τη χρήση εμπειρίας. Επομένως, για την εκπαίδευση δεν χρειάζονται δεδομένα στα οποία να έχουν επισημανθεί τα σωστά αποτελέσματα.

Προβλήματα Μη-Επιβλεπόμενης μάθησης είναι η ομαδοποίηση (ή συσταδοποίηση) και η μείωση των διαστάσεων. Πιο συγκεκριμένα:

1. Συσταδοποίηση (Clustering): όταν το σύνολο δεδομένων πρέπει να διαχωριστεί σε ομάδες, οι οποίες δεν είναι γνωστές εκ των προτέρων, αλλά δημιουργούνται κατά την εκπαίδευση του μοντέλου.
2. Μείωση διαστατικότητας (Dimensionality Reduction): όταν έχουμε σύμπτυξη δεδομένων με την αφαίρεση μεταβλητών χωρίς όμως να χάνεται η αρχική πληροφορία των δεδομένων. Αυτό μπορεί να οδηγήσει σε ευκολότερη αποθήκευση δεδομένων και ταχύτερη εκτέλεση υπολογισμών.

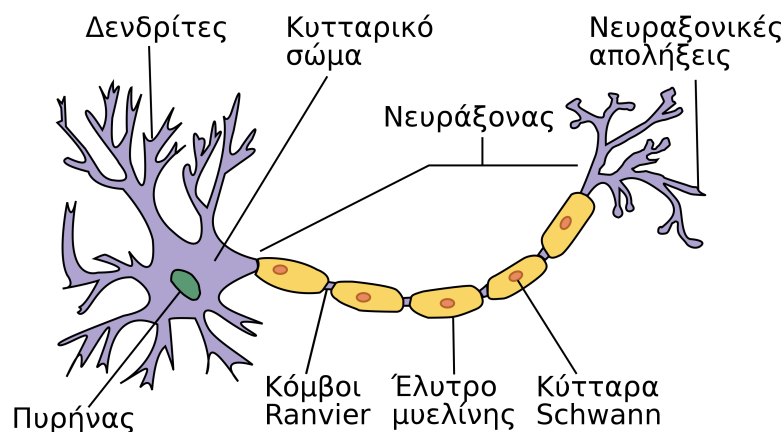
### 2.1.3 Ενισχυτική Μάθηση (Reinforcement Learning)

Οι αλγόριθμοι μαθαίνουν μια αντιστοιχίση εισόδου-εξόδου μέσω της συνεχούς αλληλεπίδρασης τους με το περιβάλλον, με στόχο το σύστημα να μεγιστοποιήσει μια συνάρτηση, την ανταμοιβή. Ο πράκτορας είναι η οντότητα που μαθαίνει και οτιδήποτε άλλο διαφορετικό από αυτό αποτελεί μέρος του περιβάλλοντος. Το περιβάλλον επιστρέφει ανταμοιβές στον πράκτορα και αυτός επιλέγει ενέργειες. Ο πράκτορας μαθαίνει από τις προηγούμενες εμπειρίες του

κάτι που μακροπρόθεσμα του επιτρέπει να επιλέγει βέλτιστα τις δράσεις. Αυτού του είδους η μάθηση συναντάται σε προβλήματα σχεδιασμού, όπως ο έλεγχος κίνησης ρομπότι, η οδήγηση ενός οχήματος, η αντιμετώπιση των αντιπάλων σε ηλεκτρονικά παιχνίδια [46].

## 2.2 Τεχνητά Νευρωνικά Δίκτυα

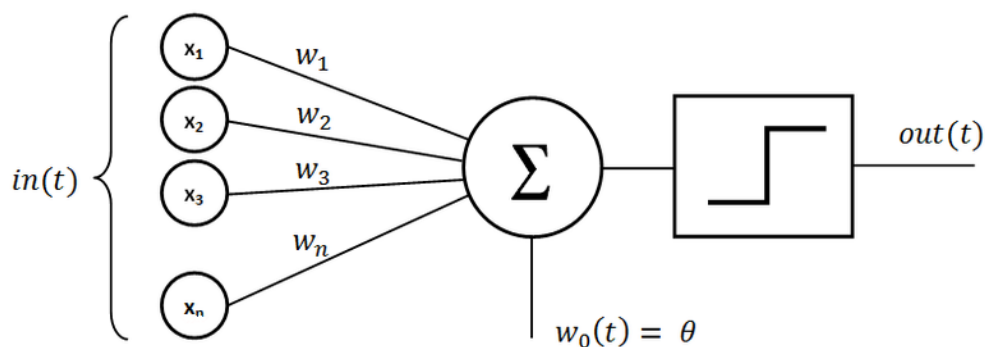
Ο καλύτερος τρόπος για να γίνει κατανοητό το πώς λειτουργεί ένα τεχνητό νευρωνικό δίκτυο είναι να κατανοηθεί πρώτα πώς λειτουργεί ένα φυσικός νευρωνικό δίκτυο μέσα στον εγκέφαλο. Το νευρωνικό δίκτυο είναι ένα κύκλωμα διασυνδεδεμένων νευρώνων το οποίο αποτελεί τμήμα νευρικού ιστού. Οι νευρώνες είναι 'η μονάδα επεξεργασίας' στον εγκέφαλο, οι οποίοι είναι υπεύθυνοι για τη μάθηση και τη διατήρηση της γνώσης. Κάθε νευρώνας αποτελείται από το σώμα, τους δενδρίτες με τους οποίους συλλέγει πληροφορίες από άλλους νευρώνες και τον άξονα μέσω του οποίου πυροδοτεί το σήμα σε άλλους νευρώνες (Εικόνα 2.1). Οι πληροφορίες που λαμβάνει κάθε νευρώνας αποτελούν τα σημάτα εισόδου του, ενώ το σήμα που εξάγει αποτελεί την έξοδο του.



Εικόνα 2.1: Νευρώνας Ανθρώπινου Εγκεφάλου (Πηγή [1])

Τα τεχνητά νευρωνικά δίκτυα είναι ένα μαθηματικό μοντέλο προσομοίωσης της πολύπλοκης λειτουργίας του ανθρώπινου εγκεφάλου. Αποτελούν μια κατηγορία της μηχανικής μάθησης και είναι ο πιο διαδομένος τρόπος ανάπτυξης μοντέλων μάθησης. Ο όρος αυτός πρωτοεμφανίστηκε το 1958 στο "Cornell Aeronautical Laboratory" από τον Frank Rosenblatt με την ονομασία "Perceptron" [47]. Ένα τεχνητό νευρικό δίκτυο αποτελείται κυρίως από τα παρακάτω μέρη (Εικόνα 2.2):

- Νευρώνες (Neurons): αποτελούν τη βασική μονάδα επεξεργασίας και του τεχνητού νευρωνικού δικτύου. Δέχεται δεδομένα εισόδου, εκτελεί κάποιους απλούς υπολογισμούς και μεταφέρει το αποτέλεσμα στον επόμενο νευρώνα. Υπάρχουν διάφορα επίπεδα νευρώνων: το επίπεδο εισόδου (input layer), ένα ή περισσότερα κρυφά επίπεδα (hidden layers) και το επίπεδο εξόδου (output layer).
- Συνάψεις και Βάρη (Synapses and Weights): αποτελούν τον τρόπο σύνδεσης των νευρώνων και μετάδοσης της πληροφορίας. Σε κάθε σύναψη αντιστοιχεί και ένα βάρος



Εικόνα 2.2: Δομή Τεχνητού Νευρωνικού Δικτύου (Πηγή [2])

$w_1, w_2, \dots, w_n$ , το οποίο με την εκπαίδευση του νευρωνικού δικτύου αλλάζει τιμές έτσι ώστε να πετυχαίνει τον σκοπό του.

- Πόλωση (Bias): επιτρέπουν στα βάρη να παίρνουν περισσότερες τιμές, οι οποίες αποθηκεύονται. Δημιουργείται έτσι μία πλουσιότερη αναπαράσταση των δεδομένων εισόδου μέσω των βαρών. Η πόλωση  $b$  δεν είναι αναγκαία αλλά πολύ χρήσιμη στα πιο πολύπλοκα νευρωνικά δίκτυα.
- Αθροιστής (Adder): προσθέτει τα σταθμισμένα, από τα αντίστοιχα βάρη, σήματα εισόδου του νευρώνα μαζί με την πόλωση.
- Συνάρτηση ενεργοποίησης  $f()$  (activation function): καθορίζει αν θα "ενεργοποιηθεί" ο νευρώνας ή όχι, δηλαδή αν η έξοδος του θα είναι 1 (ενεργός) ή 0 (ανεργός). Οι πιο γνωστές συναρτήσεις ενεργοποίησης είναι η γραμμική, η σιγμοειδής και η υπερβολική εφαπτομένη. [48, 49] Η ReLU (Rectified Linear Unit) είναι μια γραμμική συνάρτηση ενεργοποίησης που χρησιμοποιείται ευρέως και εφαρμόζει αυτή τη γραμμικότητα στις θετικές εισόδους ενώ τις αρνητικές εισόδους τις μηδενίζει. Μαθηματικά ορίζεται από την Εξίσωση 2.1 ως:

$$y = \max(0, x) \quad (2.1)$$

Η σιγμοειδής συνάρτηση έχει μια χαρακτηριστική καμπύλη σχήματος "S" και μετατρέπει την είσοδο στο διάστημα  $(0, 1)$ . Ορίζεται από την Εξίσωση 2.2

$$S(x) = \frac{1}{1 + e^{-x}} \quad (2.2)$$

Η υπερβολική εφαπτομένη είναι πολύ παρόμοια με τη σιγμοειδή αλλά έχει τη διαφορά ότι μετατρέπει την είσοδο στο διάστημα  $(-1, 1)$ . Η μαθηματική σχέση σε αυτήν την περίπτωση είναι η Εξίσωση 2.3

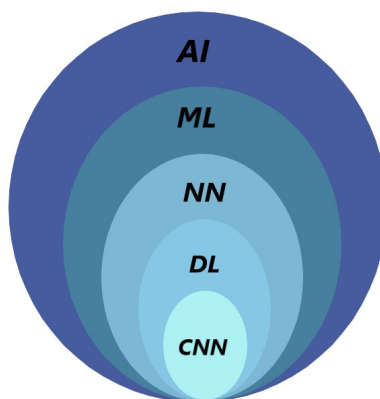
$$\frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.3)$$

Τέλος, η συνάρτηση softmax χρησιμοποιείται ως συνάρτηση ενεργοποίησης στο επίπεδο εξόδου των μοντέλων νευρωνικών δικτύων σε προβλήματα ταξινόμησης πολλών κλάσεων. Επιτρέπει την πιθανότητα κάθε τάξης και ορίζεται μαθηματικά όπως φαίνεται στην Εξίσωση 2.4:

$$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}} \quad (2.4)$$

## 2.3 Βαθιά Νευρωνικά Δίκτυα

Τα Βαθιά Νευρωνικά Δίκτυα και η Βαθιά Μάθηση, αναφέρονται σε δίκτυα που αποτελούνται από δύο ή παραπάνω κρυφά επίπεδα. Το πλεονέκτημα της βαθιάς μάθησης έναντι της παραδοσιακής μηχανικής μάθησης είναι η ικανότητα εξαγωγής χαρακτηριστικών από ανεπεξέργαστα δεδομένα [50]. Αυτό σημαίνει ότι η διαδικασία δημιουργίας χαρακτηριστικών γίνεται αυτόματα από το δίκτυο και όχι από τον προγραμματιστή. Η βαθιά μάθηση μειώνει γενικά κατά πολύ τον ανθρώπινο παράγοντα και μπορεί να εξάγει χρήσιμες πληροφορίες από μεγάλη συλλογή δεδομένων, χωρίς τη γνώση ύπαρξής τους, όπως στη μη επιβλεπόμενη μάθηση. Το μειονέκτημα της βαθιάς μάθησης είναι η απαίτηση για μεγάλο όγκο δεδομένων, πολύ χρόνο και μεγάλη υπολογιστική ισχύ. Αυτά όμως γίνονται όλο και πιο εφικτά με την εξέλιξη των υπολογιστικών συστημάτων και του διαδικτύου.



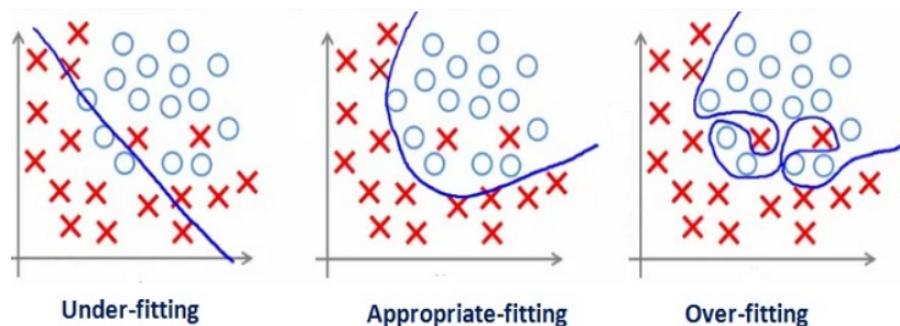
**Εικόνα 2.3:** Διάγραμμα που αναπαριστά τη σχέση μεταξύ της τεχνητής νοημοσύνης, της μηχανικής μάθησης, των νευρωνικών δικτύων, της βαθιάς μάθησης και των συνελκτικών νευρωνικών δικτύων. Η τεχνητή νοημοσύνη είναι η πιο γενική έννοια καθώς περιλαμβάνει συστήματα που αποσκοπούν να μιμηθούν την ανθρώπινη ευφυΐα.

## 2.4 Το πρόβλημα της υπερπροσαρμογής και της υποπροσαρμογής

Η υπερπροσαρμογή (overfitting) συμβαίνει όταν ένα μοντέλο μαθαίνει πολλές λεπτομέρειες και θόρυβο στα δεδομένα εκπαίδευσης σε τέτοιο βαθμό που επηρεάζει αρνητικά την απόδοση του μοντέλου σε νέα δεδομένα. Πιο απλά, το μοντέλο αποδίδει καλά στο σύνολο

λο εκπαίδευσης αλλά κακώς στο σύνολο δοκιμής, δηλαδή το μοντέλο δεν γενικεύει σε νέα δεδομένα.

Η *υποπροσαρμογή* (underfitting) είναι ακριβώς το αντίθετο. Το μοντέλο δεν μαθαίνει αρκετά από τα δεδομένα εκπαίδευσης και δεν αποδίδει ούτε στα δεδομένα εκπαίδευσης ούτε στα δεδομένα ελέγχου. Αυτό συμβαίνει συνήθως όταν δεν υπάρχουν αρκετά δεδομένα για εκπαίδευση.

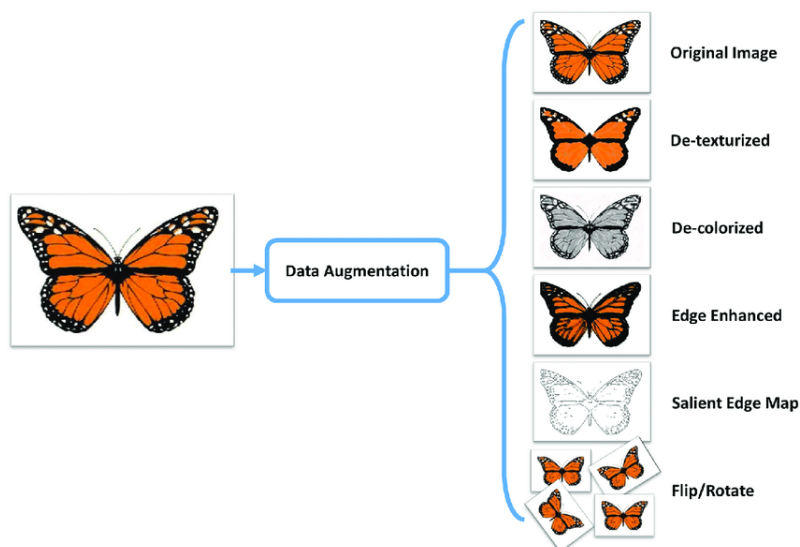


Εικόνα 2.4: Αναπαράσταση των προβλημάτων υπερπροσαρμογής και υποπροσαρμογής (Πηγή [3])

Για να εντοπιστούν φαινόμενα υπερ- και υποπροσαρμογής χωρίζουμε περαιτέρω το σύνολο δεδομένων εκπαίδευσης σε δεδομένα εκπαίδευσης και *δεδομένα επικύρωσης* (validation data), συνήθως σε αναλογία 80% και 20%, αντίστοιχα. Τα δεδομένα επικύρωσης χρησιμοποιούνται για να παρέχουν μια αμερόληπτη αξιολόγηση της προσαρμογής ενός μοντέλου στο σύνολο δεδομένων εκπαίδευσης. Επομένως, κατά την εκπαίδευσή του, το σύστημα, εκτός από την πιστότητα και την απώλεια στο σύνολο δεδομένων εκπαίδευσης, αναφέρει τα αντίστοιχα μεγέθη για το σύνολο δεδομένων επικύρωσης.

Παρακάτω επισημαίνονται τρόποι ώστε να αποφευχθεί η υπερπροσαρμογή:

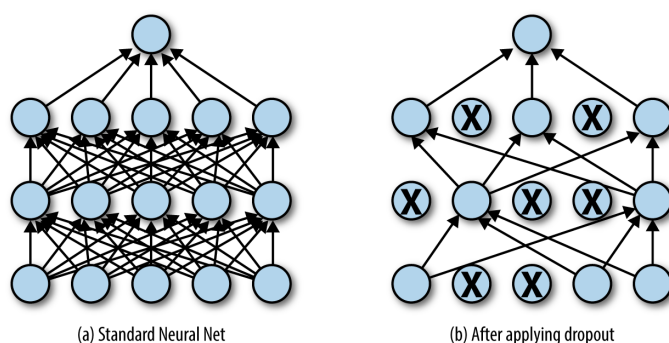
- Το πρώτο βήμα για να αντιμετωπιστεί η υπερπροσαρμογή είναι η απλοποίηση του πολύπλοκου μοντέλου που έχει χρησιμοποιηθεί, κάνοντας το μικρότερο.
- Η χρήση περισσότερων δεδομένων εκπαίδευσης. Αν δεν υπάρχουν άλλα δεδομένα μπορεί να γίνει αύξηση των υπάρχοντων δεδομένων (Data Augmentation). Μερικές από τις πιο δημοφιλείς τεχνικές είναι η αναστροφή, η περιστροφή, η κλιμάκωση, η αλλαγή φωτεινότητας, η προσθήκη θορύβου (Εικόνα 2.5).
- Η χρήση της *πρόωρου τερματισμού* (early stopping) κατά την εκπαίδευση του μοντέλου. Όταν γίνεται εκπαίδευση ενός αλγορίθμου μάθησης επαναληπτικά, μπορεί να μετρηθεί πόσο καλά αποδίδει το μοντέλο σε κάθε επανάληψη. Μέχρι έναν ορισμένο αριθμό επαναλήψεων, οι νέες επαναλήψεις βελτιώνουν το μοντέλο. Μετά από αυτό το σημείο, ωστόσο, η ικανότητα γενίκευσης του μοντέλου μπορεί να εξασθενήσει καθώς αρχίζει να προσαρμόζεται υπερβολικά στα δεδομένα εκπαίδευσης. Ο πρόωρος τερματισμός αναφέρεται στη διακοπή της εκπαιδευτικής διαδικασίας πριν ο εκπαιδευόμενος περάσει αυτό το σημείο. Μπορεί να γίνει είτε με το όρισμα του σφάλματος του δείγματος της επικύρωσης είτε με την πιστότητα του δείγματος της επικύρωσης. Δηλαδή, ψάχνουμε για μέγιστη πιστότητα και ελάχιστη απώλεια. Συχνά, το πρώτο σημάδι μη περαιτέρω



Εικόνα 2.5: Τεχνικές Data Augmentation (Πηγή [4])

βελτίωσης μπορεί να μην είναι η καλύτερη στιγμή για να σταματήσει η εκπαίδευση. Το μοντέλο μπορεί να φτάσει σε μια εποχή χωρίς βελτίωση ή και να χειροτερέψει ελαφρώς πριν γίνει πολύ καλύτερο. Επομένως, υπάρχει η δυνατότητα να προστεθεί μια καθυστέρηση στην διακοπή της εκπαίδευσης, με το όρισμα της *υπομονής* (patience), όπου δίνεται ένας ακέραιος αριθμός, που αντιστοιχεί στο πλήθος των εποχών που δεν θα έχουμε αύξηση στην πιστότητα ή μείωση στο σφάλμα.

- Το Dropout είναι μια τεχνική ομαλοποίησης που χρησιμεύει στην αντιμετώπιση της υπερπροσαρμογής των νευρωνικών δικτύων. Σε κάθε επίπεδο, μεμονωμένοι κόμβοι είτε 'αποσύρονται' από το δίκτυο με πιθανότητα  $1 - p$  ή διατηρούνται με πιθανότητα  $p$ , ώστε να προκύψει ένα απλουστευμένο δίκτυο. Οι εισερχόμενες και εξερχόμενες ακμές σε έναν κόμβο που 'αποσύρθηκε' αφαιρούνται επίσης (Εικόνα 2.6). Μόνο το προκύπτον δίκτυο εκπαιδεύεται στα δεδομένα και στα επόμενα επίπεδα οι κόμβοι εισάγονται ξανά στο δίκτυο με τα αρχικά τους βάρη [51].



Εικόνα 2.6: Τεχνική Dropout (Πηγή [5])

- Η ομαλοποίηση L1 και L2 είναι οι δύο πιο χρησιμοποιούμενοι μέθοδοι ομαλοποίησης για την αντιμετώπιση της υπερπροσαρμογής. Η ομαλοποίηση είναι ένας επιπλέον όρος που προστίθεται στη συνάρτηση απώλειας για την επιβολή ποινής σε μεγάλα

βάρη παραμέτρων δικτύου για τη μείωση της υπερπροσαρμογής. Στην ομαλοποίηση L1 προστίθεται η νόρμα L1, όπου L1 το άθροισμα των βαρών, πολλαπλασιασμένη με μια υπερπαραμέτρο  $\lambda$ . Στην L2 προστίθεται το γινόμενο της νόρμας L2, όπου L2 το τετράγωνο των βαρών, με το μισό της υπερπαραμέτρου  $\lambda$ .

## 2.5 Συνελικτικά Νευρωνικά Δίκτυα (CNNs)

### 2.5.1 Ορισμός

Τα συνελικτικά δίκτυα (Convolutional Neural Networks / CNNs / ConvNets) αποτελούν μια κατηγορία βαθιών νευρωνικών δικτύων, τα οποία είναι τα πιο διαδεδομένα μοντέλα για την ανάλυση εικόνας και την όραση υπολογιστών [52].

Τα σύγχρονα CNNs εκτός από την αποτελεσματικότητα του δείγματος στην επίτευξη μοντέλων υψηλής ακρίβειας προβλέψεων, τείνουν να είναι υπολογιστικά αποδοτικά, τόσο επειδή απαιτούν λιγότερες παραμέτρους από τις πλήρως συνδεδεμένες αρχιτεκτονικές (Fully Connected Architectures) όσο και επειδή οι συνελίξεις είναι εύκολο να παραλληλιστούν μεταξύ των πυρήνων GPU. Κατά συνέπεια, οι ερευνητές συχνά εφαρμόζουν τα συνελικτικά δίκτυα όλο και περισσότερο αφού εμφανίζονται αξιόπιστα ακόμα και σε εργασίες με μονοδιάστατη δομή ακολουθίας, όπως η ανάλυση ήχου, κειμένου και χρονοσειρών, όπου συμβατικά χρησιμοποιούνται επαναληπτικά νευρωνικά δίκτυα. Ορισμένες έξυπνες προσαρμογές των συνελικτικών δικτύων τα έφεραν επίσης σε επαφή με δεδομένα δομημένα σε γραφήματα και συστήματα προτάσεων.

Το συνελικτικό νευρωνικό δίκτυο, όπως δηλώνει και το όνομα του, προέρχεται από τη μαθηματική πράξη της συνέλιξης. Η συνέλιξη μεταξύ δύο συναρτήσεων  $f, g : R \rightarrow R$  ορίζεται από την Εξίσωση 2.5 ως:

$$(f * g)(x) = \int f(z)g(x - z)xdz \quad (2.5)$$

Μετράμε δηλαδή την επικάλυψη μεταξύ  $f$ ,  $g$  όταν η δεύτερη συνάρτηση αναστρέφεται και μετατοπίζεται κατά  $x$ . Αν έχουμε διακριτές συναρτήσεις, το ολοκλήρωμα μετατρέπεται σε άθροισμα όπως φαίνεται στην Εξίσωση 2.6:

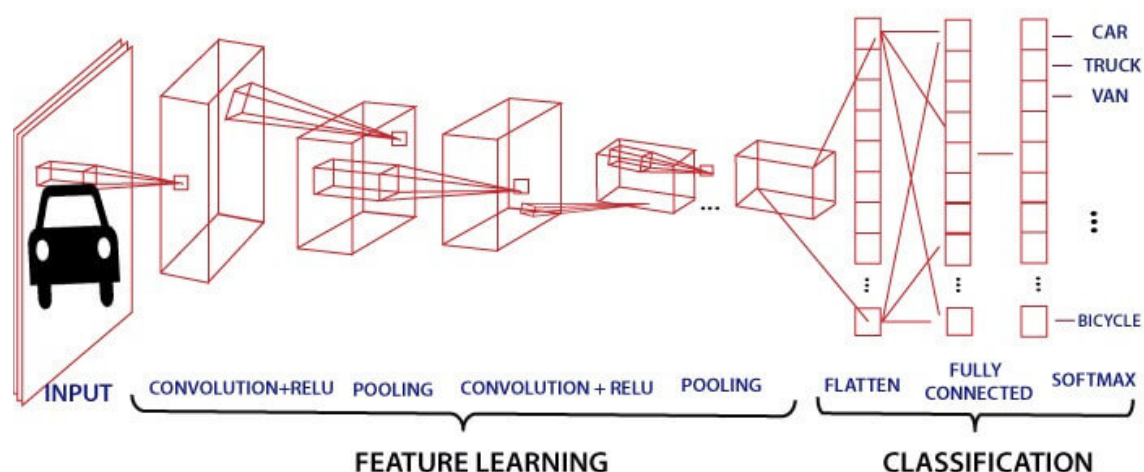
$$(f * g)(i) = \sum_a f(a)g(i + a) \quad (2.6)$$

Το συνελικτικό νευρωνικό δίκτυο αποτελείται, όπως και ένα απλό νευρωνικό δίκτυο, από το επίπεδο εισόδου, τα κρυφά επίπεδα και το επίπεδο εξόδου. Σε ένα συνελικτικό δίκτυο όμως, τα κρυφά επίπεδα περιλαμβάνουν επίπεδα που κάνουν τη μαθηματική πράξη της συνέλιξης στην οποία η διοδιάστατη είσοδος πολλαπλασιάζεται με έναν διοδιάστατο πίνακα από βάρη, που ονομάζεται *πυρήνας* ή *φίλτρο*.

Ένα συνελικτικό νευρωνικό δίκτυο μπορεί να χωριστεί σε δύο μέρη: στην εξαγωγή των χαρακτηριστικών από την είσοδο και στην ταξινόμηση και τη δημιουργία της εξόδου (Εικόνα 2.7).

Η είσοδος ενός συνελικτικού δικτύου μπορεί να είναι μια ή περισσότερες εικόνες. Η διάσταση αυτής είναι: (Πλήθος Εικόνων) X (Μήκος Εικόνας) X (Πλάτος Εικόνας) X (Αριθμός





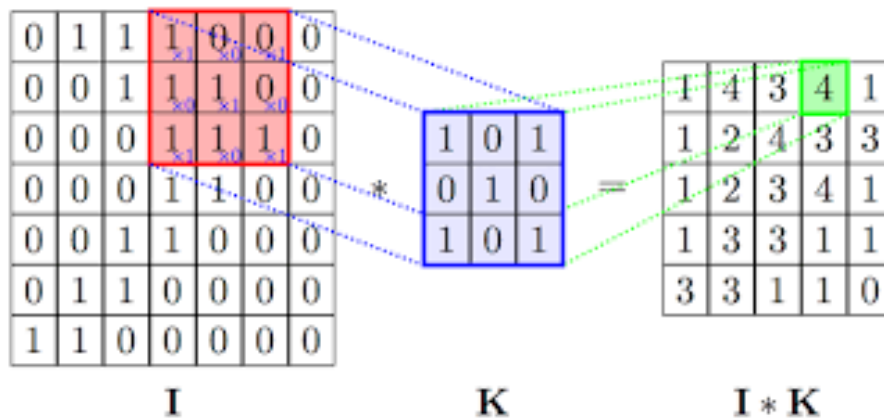
Εικόνα 2.7: Παράδειγμα συνελκτικού νευρωνικού δικτύου (Πηγή [6])

Καναλιών). Κάθε εικόνα είναι ένας πίνακας, και κάθε τιμή του πίνακα αντιπροσωπεύει την τιμή ενός εικονοστοιχείου. Στις ασπρόμαυρες εικόνες, όπου ο αριθμός των καναλιών είναι 1, οι τιμές των εικονοστοιχείων κυμαίνονται από 0 (μαύρο) έως 255 (λευκό), ή με κανονικοποίηση στο διάστημα  $[0, 1]$ . Για έγχρωμες εικόνες, χρησιμοποιείται συχνά ο χρωματικός χώρος RGB, όπου ο αριθμός των καναλιών είναι 3. Πιο συγκεκριμένα, κάθε εικονοστοιχείο αναπαρίστανται από 3 τιμές, καθε μια απο τις οποίες αναφέρεται στο επίπεδο κόκκινου, πρασίνου και μπλε.

Η εξαγωγή των χαρακτηριστικών μιας εικόνας, πραγματοποιείται στα κρυφά επίπεδα. Η διαδικασία αυτή πραγματοποιείται σταδιακά. Στα πρώτα επίπεδα ενός συνελκτικού δικτύου, εξάγονται χαρακτηριστικά χαμηλού επιπέδου, όπως οι ακμές. Στα επόμενα επίπεδα, τα χαρακτηριστικά αυτά συντίθεται δημιουργώντας πιο σύνθετες αναπαραστάσεις (μέρη αντικειμένων). Όσο πιο βαθύ γίνεται το δίκτυο τα μέρη αυτά ομαδοποιούνται και δημιουργούν τα τελικά αντικείμενα.

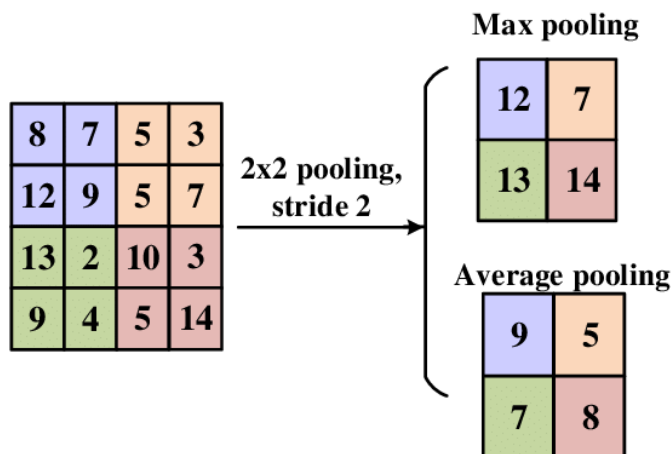
## 2.5.2 Επίπεδα Συνελκτικών Νευρωνικών Δικτύων

- Συνελκτικό Επίπεδο (Convolutional Layer): Σε εφαρμογές επεξεργασίας εικόνας, χρησιμοποιούμε τους όρους *πυρήνα* (kernel) ή *φίλτρο* (filter). Ως πυρήνας μπορεί να οριστεί ένας πίνακας διαστάσεων: (Μήκος Φίλτρου) X (Πλάτος Φίλτρου) X (Πλήθος Καναλιών εικόνας). Τα στοιχεία του πυρήνα πολλαπλασιάζονται με ένα τμήμα της εικόνας (το οποίο πρέπει να έχει τις ίδιες διαστάσεις) και τα αποτελέσματα τους τοποθετούνται στον πίνακα εξόδου, ο οποίος ονομάζεται *χάρτης ενεργοποίησης* (activation map) ή *χάρτης χαρακτηριστικών* (feature map) (Εικόνα 2.8). Η τιμή του χάρτη σε κάθε θέση εκφράζει την πιθανότητα να βρίσκεται εκεί το επιθυμητό χαρακτηριστικό. Επιπλέον, ο πυρήνας μπορεί να ολισθαίνει σε κάθε μετακίνησή του, με βήμα (stride) ίσο ή μεγαλύτερο του 1. Σημειώνεται ότι είναι δυνατόν να έχουμε έξοδο πολλούς χάρτες χαρακτηριστικών με χρήση πολλών πυρήνων ώστε να εξάγουμε αρκετά χαρακτηριστικά από μια εικόνα.



Εικόνα 2.8: Εφαρμογή συνέλιξης φίλτρου σε εικόνα εισόδου (Πηγή [7])

- Επίπεδο Ενεργοποίησης (Activation Layer): Τα περισσότερα συστήματα στα οποία εφαρμόζονται τα συνελκτικά δίκτυα CNN δεν είναι γραμμικά. Για να εισαχθεί μη γραμμικότητα απαιτείται ένα επίπεδο ενεργοποίησης μετά από κάθε συνελκτικό επίπεδο, το οποίο εφαρμόζει στη έξοδό του μία συνάρτηση ενεργοποίησης. Η πιο διαδεδομένη συνάρτηση ενεργοποίησης για αυτό το σκοπό είναι η Rectified Linear Unit (ReLU).



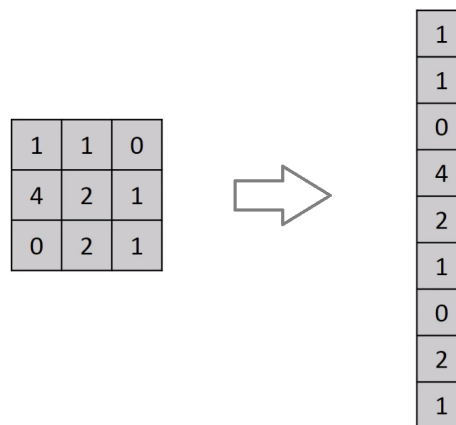
Εικόνα 2.9: Παράδειγμα υποδειγματοληψίας μεγίστου και μέσου όρου (Πηγή [8])

- Επίπεδο Υποδειγματοληψίας (Pooling Layer): Μπορεί να παρεμβάλλεται μεταξύ των συνελκτικών επιπέδων και σκοπεύει στην μείωση των διαστάσεων των χαρτών ενεργοποίησης. Καθορίζεται το μέγεθος του παραθύρου στο οποίο θα εφαρμοστεί υποδειγματοληψία και τα πιο συνηθισμένα είδη είναι η υποδειγματοληψία μεγίστου (max pooling), η υποδειγματοληψία μέσου όρου (average pooling) ή η στοχαστική υποδειγματοληψία (Εικόνα 2.9). Εκτός από τη βελτίωση της ταχύτητας εκπαίδευσης, η ύπαρξη του επιπέδου αυτού ελαχιστοποιεί και την πιθανότητα υπερεκπαίδευσης του δικτύου.
- Επίπεδο Κανονικοποίησης Δέσμης (Batch Normalization Layer): Η δομή των Συνελκτικών Νευρωνικών Δικτύων με τα πολλά διαδοχικά επίπεδα, συχνά οδηγεί σε ένα φαινόμενο που ονομάζεται “internal covariate shift”, κατά το οποίο η προσαρμογή των παραμέτρων του δικτύου κατά την εκπαίδευση προκαλεί αλλαγή στην κατανομή των

ενεργοποιήσεων των διαφόρων επιπέδων (συνήθως των τελευταίων). Για να αποφευχθεί αυτό το φαινόμενο, συνήθως προσθέτουμε επίπεδα κανονικοποίησης δέσμης, τα οποία εξασφαλίζουν την κανονικοποίηση των δεδομένων κάθε δέσμης σε κάθε επίπεδο. Η κανονικοποίηση γίνεται με χρήση των στατιστικών χαρακτηριστικών του υποσυνόλου, ώστε ο μέσος όρος να ισούται με 0 και η διακύμανση να είναι 1.

Ο τελικός στόχος ενός συνελκτικού νευρωνικού δικτύου είναι η ταξινόμηση των εικόνων σύμφωνα με την εργασία. Αυτό μπορεί να επιτευχθεί με την εισαγωγή των τελικών εξαγόμενων χαρακτηριστικών σε έναν ταξινομητή. Η ταξινόμηση (Classification) γίνεται στα παρακάτω επίπεδα.

- Flatten Layer: Οι τιμές των τελικών χαρτών ενεργοποίησης, όπως προέκυψαν μετά από διαδοχικά συνελκτικά επίπεδα, αποτελούν την είσοδο του πλήρως συνδεδεμένου δικτύου. Όμως, αυτό το επίπεδο, αναμένει στην είσοδό του ένα μονοδιάστατο διάνυσμα αριθμών, γι' αυτό και η τρισδιάστατη έξοδος του προηγούμενου επιπέδου μετατρέπεται σε μονοδιάστατη με το λεγόμενο flatten layer (Εικόνα 2.10).



Εικόνα 2.10: Παράδειγμα μετατροπής σε μονοδιάστατη έξοδο (Πηγή [9])

- Πλήρως Συνδεδεμένο Επίπεδο (Fully Connected Layer - FC): Αποτελεί ένα επίπεδο οι νευρώνες του οποίου συνδέονται με όλους τους νευρώνες του προηγούμενου επιπέδου. Η είσοδος ενός πλήρως συνδεδεμένου στρώματος διέρχεται μέσω της συνάρτησης ενεργοποίησης που χρησιμοποιείται, προκειμένου να αποκτηθεί η έξοδος. Ένα τέτοιο επίπεδο, εξάγει την τελική έξοδο του δικτύου στα προβλήματα ταξινόμησης και παλινδρόμησης [53].

### 2.5.3 Παράμετροι και Διαστάσεις Εξόδων

Παρακάτω εξηγείται πως προκύπτει ο συνολικός αριθμός παραμέτρων που μαθαίνει το δίκτυο σε κάθε επίπεδο κατά τη φάση της εκπαίδευσης. Μαθηματικά το πλήθος των παραμέτρων σε κάθε επίπεδο φαίνεται στις Εξισώσεις 2.7-2.9.

- Επίπεδο εισόδου (Input Layer): Το στρώμα εισόδου δεν έχει παραμέτρους, αυτό που κάνει είναι απλώς να παρέχει τις διαστάσεις της εικόνας εισόδου.

- Συνελικτικό επίπεδο (Convolutional Layer): Εδώ μαθαίνει το CNN, οπότε έχουμε πίνακες με βάρη. Οι παράμετροι υπολογίζονται ως εξής: Αρχικά, υπολογίζονται οι παράμετροι που προκύπτουν με το ένα φίλτρο. Πολλαπλασιάζεται το πλάτος του φίλτρου  $m$  με το ύψος του φίλτρου  $h$  και το 3, αν είναι έγχρωμη η είσοδος ή 1 αν είναι ασπρόμαυρη ( $d$ ), και προστίθεται ο όρος μεροληψίας. Το αποτέλεσμα αυτό πολλαπλασιάζεται με τον αριθμό των φίλτρων  $k$  του τρέχοντος επιπέδου. Άρα ο συνολικός αριθμός των παραμέτρων σε ένα επίπεδο CONV είναι:

$$(m \cdot h \cdot d + 1) \cdot k \quad (2.7)$$

,όπου 1 ο όρος πόλωσης για κάθε φίλτρο.

- Επίπεδο Υποδειγματοληψίας (Pooling Layer): Το επίπεδο αυτό δεν έχει παραμέτρους για εκμάθηση, αφού το μόνο που κάνει είναι να υπολογίζει έναν συγκεκριμένο αριθμό, χωρίς να εμπλέκεται μάθηση backpropagation.
- Πλήρως συνδεδεμένο επίπεδο (FC): Αυτό το επίπεδο έχει το μεγαλύτερο πλήθος παραμέτρων εκμάθησης. Ο υπολογισμός των παραμέτρων προκύπτει από το άθροισμα του γινόμενου του αριθμού των νευρώνων στο τρέχον επίπεδο  $c$  και του αριθμού των νευρώνων στο προηγούμενο στρώμα  $p$  και τον όρο μεροληψίας. Άρα, ο αριθμός των παραμέτρων είναι:

$$c \cdot p + 1 \cdot c \quad (2.8)$$

- Επίπεδο Κανονικοποίησης δέσμης (Batch Normalization Layer): Αυτό το επίπεδο έχει παραμέτρους. Πιο συγκεκριμένα έχει 2 παραμέτρους που μαθαίνονται  $\beta$  και  $\gamma$  και 2 μη-εκπαιδευσιμες παραμέτρους (Mean Moving Average και Variance Moving Average). Αυτές οι 4 παράμετροι πολλαπλασιάζονται με το πλήθος των φίλτρων  $k$  του τρέχοντος επιπέδου. Άρα, ο αριθμός των παραμέτρων είναι:

$$4 \cdot k \quad (2.9)$$

- Επίπεδο Dropout (Dropout Layer): Το επίπεδο αυτό δεν έχει παραμέτρους.

Τώρα, θα εξηγηθεί πως υπολογίζονται οι διαστάσεις των εξόδων μιας εικόνας όταν διέρχεται από ένα συνελικτικό επίπεδο, από ένα επίπεδο υποδειγματοληψίας, από ένα επίπεδο “flatten” και τέλος ένα πλήρως συνδεδεμένο επίπεδο. Μαθηματικά, ορίζονται απο τις Εξισώσεις 2.10-2.11. Οι παράμετροι που επηρεάζουν το σχήμα εξόδου είναι:

1. Οι διαστάσεις της εικόνας εισόδου:  $(i \times i)$
2. Το μέγεθος του φίλτρου/πυρήνα:  $(f \times f)$
3. Βήματα (strides):  $S$  (ακέραιος αριθμός)
4. Padding:  $P$  (ακέραιος)

5. Βάθος χαρτών χαρακτηριστικών:  $D$  (ακέραιος αριθμός)

- Διάσταση εξόδου συνέλιξης:

$$\left(\frac{i-f+2 \cdot P}{S}+1\right) \times D \quad (2.10)$$

- Διάσταση εξόδου ενός Pooling Layer:

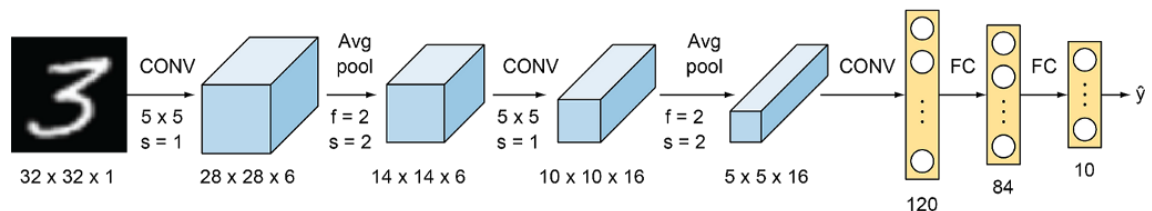
$$\left(\frac{i-f}{S}+1\right) \times D \quad (2.11)$$

- Το επίπεδο Flatten χρησιμοποιείται για να κάνει την πολυδιάστατη είσοδο μονοδιάστατη, που χρησιμοποιείται στη μετάβαση από το επίπεδο συνέλιξης στο πλήρως συνδεδεμένο επίπεδο. Άρα, η έξοδος από αυτό το επίπεδο θα έχει μια διάσταση η οποία είναι αποτέλεσμα του γινομένου των διαστάσεων της εισόδου του.
- Η έξοδος του πλήρως-διασυνδεδεμένου επιπέδου θα επηρεαστεί από τον αριθμό των νευρώνων που καθορίζονται. Για παράδειγμα, εάν η είσοδος είναι (8,) και οι νευρώνες είναι 16, τότε το σχήμα εξόδου είναι (16,).

#### 2.5.4 Η Εξέλιξη των Συνελκτικών Νευρωνικών Δικτύων

Τα συνελκτικά νευρωνικά δίκτυα παρουσιάστηκαν πρώτη φορά το 1989 από τον Yann LeCun, έναν μεταδιδακτορικό ερευνητή επιστήμης υπολογιστών. Ο LeCun βασίστηκε στο “neocognitron” [54] ένα βασικό νευρωνικό δίκτυο αναγνώρισης εικόνας, που εφηύρε ο Kunihiko Fukushima, ένας Ιάπωνας επιστήμονας το 1979.

Η πρώτη έκδοση των CNN, που ονομάστηκε LeNet (από τον LeCun), μπορούσε να αναγνωρίσει χειρόγραφα ψηφία (Εικόνα 2.11). Πιο συγκεκριμένα, οι LeCun et al. εκπαίδευσαν ένα CNN 7 επιπέδων, χρησιμοποιώντας τη βάση δεδομένων MNIST [55], που περιλαμβάνει εικόνες  $28 \times 28$  χειρόγραφων ψηφίων σε συνδυασμό με την πραγματική τους ετικέτα (0, 1, 2, 3, 4, 5, 6, 7, 8 ή 9). Τα CNN βρήκαν εξειδικευμένη εφαρμογή στις τραπεζικές και ταχυδρομικές υπηρεσίες και στον τραπεζικό τομέα, όπου αναγνώριζαν ταχυδρομικούς κώδικες σε φακέλους και ψηφία στις επιταγές. [56]

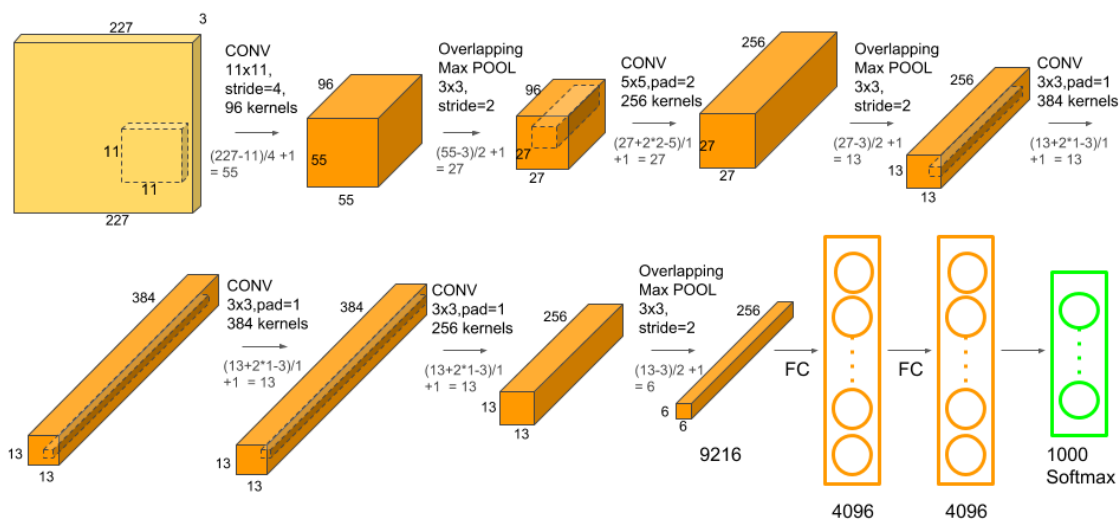


Εικόνα 2.11: Αρχιτεκτονική LeNet (Πηγή [10])

Τα συνελκτικά νευρωνικά δίκτυα, παρά την εφευρετικότητάς τους, παρέμειναν στο περιθώριο της όρασης υπολογιστών και της τεχνητής νοημοσύνης. Εκείνη την εποχή, η τεχνική δεν ήταν εύκολα εφαρμόσιμη σε εικόνες με υψηλή ανάλυση αφού χρειαζόνταν πολλά δεδομένα και υπολογιστικούς πόρους για να λειτουργούν αποτελεσματικά.

Το 2004, αποδείχθηκε από τους K. S. Oh και K. Jung ότι τα τυπικά νευρωνικά δίκτυα μπορούν να επιταχυνθούν σε μεγάλο βαθμό με χρήση GPU. Η υλοποίησή τους ήταν 20 φορές ταχύτερη από μια αντίστοιχη υλοποίηση σε CPU [57, 58]. Η πρώτη GPU-υλοποίηση ενός CNN περιγράφηκε το 2006 από τους K. Chellapilla et al [59]. Η υλοποίησή τους ήταν 4 φορές ταχύτερη από μια αντίστοιχη εφαρμογή σε CPU.

Το 2012, τα CNNs απέκτησαν μεγάλη δημοτικότητα όταν ο Alex Krizhevsky δημοσίευσε μια εργασία στην οποία παρουσίασε την αρχιτεκτονική AlexNet η οποία κέρδισε τον διαγωνισμό “ImageNet Large Scale Visual Recognition Challenge” (ILSVRC). Το AlexNet περιέχει πέντε επίπεδα συνέλιξης, τρία επίπεδα υποδειγματοληψίας μεγίστου, τρία πλήρως συνδεδεμένα επίπεδα και δύο επίπεδα κανονικοποίησης (Εικόνα 2.12). Αυτή η βαθιά αρχιτεκτονική των 60 εκατομμυρίων παραμέτρων εκπαιδεύτηκε χρησιμοποιώντας την Compute Unified Device Architecture (CUDA) σε δύο GTX 580 GPU [60].

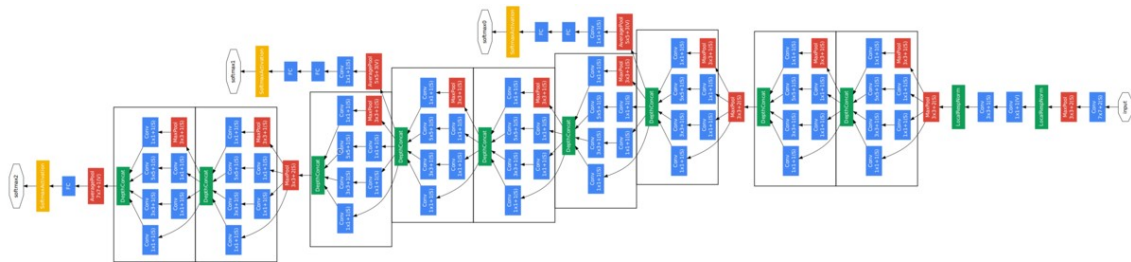


Εικόνα 2.12: Αρχιτεκτονική AlexNet (Πηγή [11])

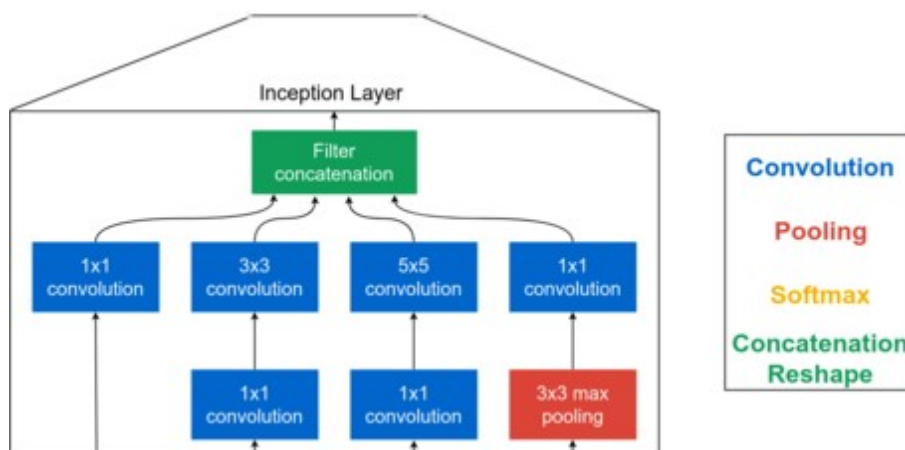
Η διαθεσιμότητα μεγάλων συνόλων δεδομένων, συγκεκριμένα του συνόλου δεδομένων ImageNet με εκατομμύρια ετικέτες εικόνων και τεράστιων υπολογιστικών πόρων, επέτρεψε στους ερευνητές να δημιουργήσουν πολύπλοκα CNN που θα μπορούσαν να εκτελούν εργασίες όρασης υπολογιστή που προηγουμένως ήταν αδύνατες. Το ImageNet [61] είναι ένα δημόσιο, ελεύθερα διαθέσιμο σύνολο δεδομένων εικόνων και των αντίστοιχων ετικετών τους. Αντί να εστιάζει σε χειρόγραφα ψηφία με την ένδειξη 0-9, το ImageNet εστιάζει σε φυσικές εικόνες. Οι ετικέτες των εικόνων αποκτήθηκαν μέσω τεράστιας ανθρώπινης προσπάθειας (με χειροκίνητη επισήμανση).

Στο ILSVRC 2014, σχεδόν όλες οι διαγωνιζόμενες ομάδες χρησιμοποίησαν CNN. Ο νικητής ήταν η ομάδα της Google που παρουσίασε το GoogLeNet[62] που είναι 22 επιπέδων. Η προτεινόμενη αρχιτεκτονική αύξησε τη μέση ακρίβεια ανίχνευσης αντικειμένων στο 0,439329 και μείωσε το σφάλμα ταξινόμησης στο 0,06656, το καλύτερο αποτέλεσμα μέχρι σήμερα. Το GoogLeNet ή αλλιώς InceptionV1 είναι η πρώτη αρχιτεκτονική που χρησιμοποίησε τη μονάδα

“inception”. Αυτή η μονάδα βασίστηκε στην ιδέα ύπαρξης φίλτρων πολλαπλών μεγεθών που μπορούν να λειτουργούν στο ίδιο επίπεδο. Πιο αναλυτικά, η λειτουργία συνέλιξης εκτελείται σε εισόδους με τρία μεγέθη φίλτρων: (1x1), (3x3) και (5x5). Μια λειτουργία υποδειγματοληψίας μεγίστου εκτελείται επίσης με τις συνελιξεις και στη συνέχεια στέλνεται στην επόμενη μονάδα inception. Αυτό επιτυγχάνει τον περιορισμό του αριθμού των καναλιών εισόδου και τη μείωση των διαστάσεων του δικτύου, ώστε να εκτελεί ταχύτερους υπολογισμούς. Έτσι, το δίκτυο καταφέρνει και γίνεται ευρύτερο και όχι βαθύτερο.



Εικόνα 2.13: Αρχιτεκτονική GoogleNet (Πηγή [12])

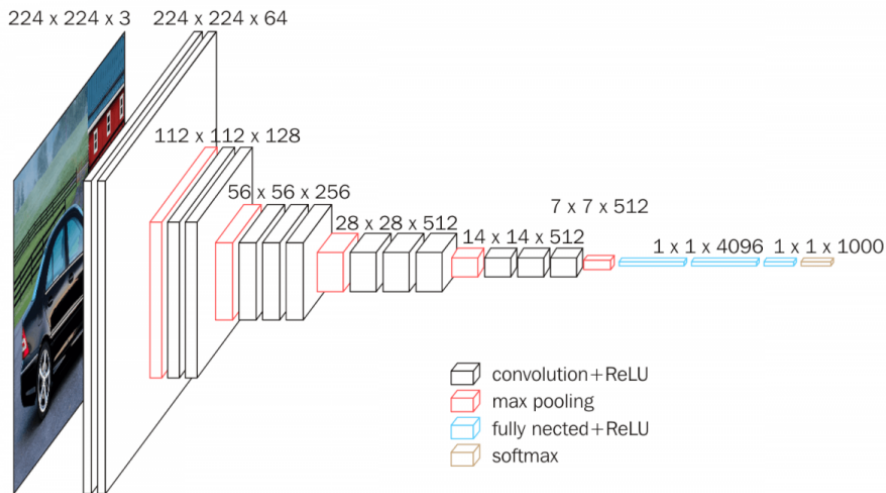


Εικόνα 2.14: Inception μονάδα της GoogleNet (Πηγή [12])

Παραπάνω, φαίνεται σχηματικά το μοντέλο GoogleNet (Εικόνα 2.13). Σε τετράγωνα κουτιά βρίσκονται οι μονάδες inception, τα μπλε κουτιά είναι οι συνελιξεις, τα κόκκινα κουτιά αφορούν τα επίπεδα μέγιστης υποδειγματοληψίας, τα κίτρινα είναι τα επίπεδα ενεργοποίησης softmax και τα πράσινα είναι τα φίλτρα συνένωσης. Από κάτω δίνεται σχηματικά και σε καλύτερη ανάλυση η μονάδα inception (Εικόνα 2.14).

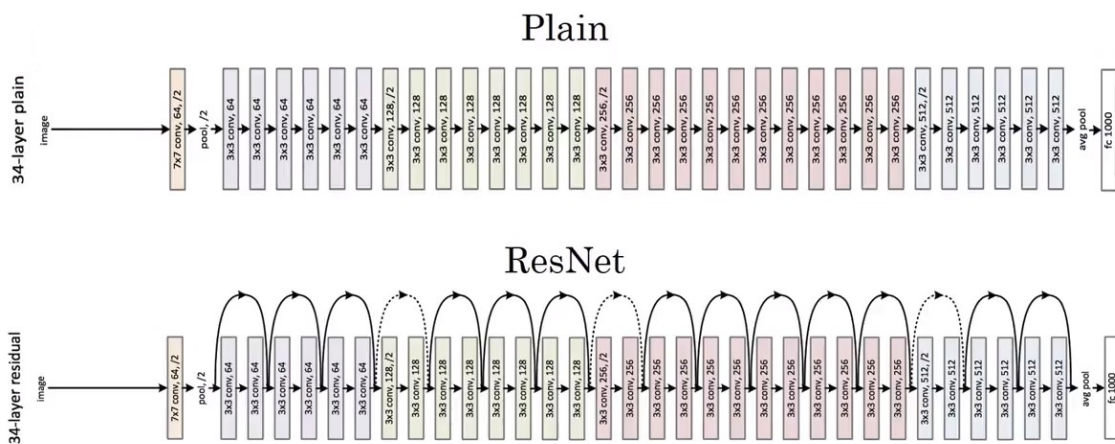
Το VGG16 [63] είναι ένα άλλο διάσημο μοντέλο CNN που πήρε μέρος στο διαγωνισμό ILSVRC-2014 και πήρε τη δεύτερη θέση. Προτάθηκε από τους Karen Simonyan και Andrew Zisserman από το Πανεπιστήμιο της Οξφόρδης και επιτυγχάνει 92,7% ακρίβεια δοκιμών top-5 στο ImageNet. Είναι μια βελτίωση σε σχέση με το AlexNet αντικαθιστώντας τα μεγάλα φίλτρα μεγέθους πυρήνα (11 και 5 στο πρώτο και δεύτερο συνελκτικό επίπεδο, αντίστοιχα) με πολλαπλά φίλτρα μεγέθους πυρήνα  $3 \times 3$  το ένα μετά το άλλο (Εικόνα 2.15). Το VGG16 εκπαιδεύτηκε για εβδομάδες και χρησιμοποιούσε GPU NVIDIA Titan Black.

Η αρχιτεκτονική ResNet (Residual Network) [64] ήταν η νικήτρια του ILSVRC 2015 χρησιμοποιώντας 122 επίπεδα. Η ResNet έχει πολλές παραλλαγές: ResNet-18, ResNet-34,



Εικόνα 2.15: Αρχιτεκτονική VGG16 (Πηγή [13])

ResNet-50, ResNet-101, ResNet-110, ResNet-152, ResNet-164, ResNet-1202 κ.α. Η διαφορά σε αυτές είναι μόνο στο πλήθος των επιπέδων τους. Η αρχιτεκτονική αυτή προτάθηκε επί αφορμή του τεχνικού προβλήματος της εξαφανιζόμενης κλίσης (vanishing gradient). Η εξαφανιζόμενη κλίση συμβαίνει όταν αυξηθεί πολύ ο αριθμός των επιπέδων, με αποτέλεσμα η επίδοση του δικτύου να μειωθεί λόγω της δυσκολίας βελτιστοποίησης με οπισθοδιάδοση. Το ζήτημα αυτό αντιμετωπίστηκε με τη μονάδα “residual block”. Το residual block περιέχει δύο σειριακά συνδεδεμένα επίπεδα συνέλιξης και μια σύνδεση “skip connection” που αθροίζει την είσοδο στην έξοδο πριν το επίπεδο ενεργοποίησης (Εικόνα 2.16). Αυτή η σύνδεση διευκολύνει την διαδικασία μάθησης, ενώ προσφέρει καλύτερη κατανομή της κλίσης σφάλματος προς τα πρώτα επίπεδα.



Εικόνα 2.16: Αντιπαράθεση ενός απλού δικτύου 34 επιπέδων και ενός residual δικτύου (Πηγή [14])



## Κεφάλαιο **3**

# Σχετικές εργασίες και περιγραφή επιλεγμένων αρχιτεκτονικών

---

**Σ**ε αυτό το κεφάλαιο θα αναφερθούν αρκετές δημοσιευμένες, σχετικές μελέτες που έχουν γίνει στην ανίχνευση φωτιών από εικόνες και βίντεο. Θα παρουσιαστούν αρχικά μελέτες με παλαιότερες μεθόδους και μετέπειτα πιο σύγχρονες προσεγγίσεις. Αφορμώμενοι από αυτές θα εστιάσουμε στις ‘καλύτερες’ για το συγκεκριμένο πρόβλημα και θα αναπτύξουμε αναλόγως τα δικά μας μοντέλα.

### 3.1 Σχετικές εργασίες

#### 3.1.1 Κλασικές προσεγγίσεις

Η μια ‘ομάδα’ παραδοσιακών προσεγγίσεων σε αυτόν τον τομέα επικεντρώνεται στη προσέγγιση που βασίζεται αποκλειστικά στο χρώμα. Στο άρθρο [65] έχουμε μια προσέγγιση που θεωρεί μια τιμή κατωφλίου χρώματος. Παρουσιάζεται ένα σύστημα που παίρνει ως είσοδο έγχρωμο βίντεο και μελετώνται σε αυτό οι φασματικές, χωρικές και χρονικές ιδιότητες της πυρκαγιάς. Στην εργασία [66] υιοθετείται ένα μοντέλο RGB που βασίζεται στη χρωματική μέτρηση και τη μέτρηση της ‘αταξίας’ για την εξαγωγή εικονοστοιχείων πυρκαγιάς και εικονοστοιχείων καπνού. Η συνάρτηση απόφασης των εικονοστοιχείων φωτιάς συνάγεται κυρίως από την ένταση και τον κορεσμό της συνιστώσας R. Τα εξαγόμενα εικονοστοιχεία πυρκαγιάς θα επαληθευτούν εάν πρόκειται για πραγματική φωτιά τόσο από τη δυναμική της ανάπτυξης και την αταξία όσο και από περαιτέρω καπνό. Αν με επαναληπτικό έλεγχο, ο λόγος των φλογών αυξάνεται, δίνεται συναγερμός πυρκαγιάς, εφόσον πληρείται η συνθήκη συναγερμού.

Στο [67] χρησιμοποιείται ο εναλλακτικός χρωματικός χώρος YCbCr για την κατασκευή ενός γενικού μοντέλου χρωματισμού για ταξινόμηση εικονοστοιχείων φλόγας. Το προτεινόμενο μοντέλο δίνει 99,0% σωστό ρυθμό ταξινόμησης εικονοστοιχείων φλόγας με ποσοστό ψευδούς συναγερμού 31,5%. Κάθε εικονοστοιχείο σε μια έγχρωμη εικόνα που περιέχει μια πυρκαγιά, η τιμή του κόκκινου καναλιού είναι μεγαλύτερη από το πράσινο κανάλι και η τιμή του πράσινου καναλιού είναι μεγαλύτερη από την τιμή του μπλε καναλιού για τη χωρική θέση. Επιπλέον, το χρώμα της φλόγας έχει υψηλό κορεσμό στο κόκκινο κανάλι. Στο άρθρο [68] παρουσιάζεται μια ολοκληρωμένη ανασκόπηση ανίχνευσης καπνού και πυρκαγιάς χρησιμοποιώντας επεξεργασία εικόνας. Ο καπνός είναι ένας καλός δείκτης μιας κατάστα-

σης πριν από την πυρκαγιά και πολλές πυρκαγιές είναι δείκτες επακόλουθων επικίνδυνων καταστάσεων λόγω της εξάπλωσης της φωτιάς. Πρώτα γίνεται σύγκριση μεθόδων ανίχνευσης καπνού και διαφορετικών τύπων προσεγγίσεων για την ταξινόμηση του καπνού. Επιπλέον, αναλύονται διαφορετικοί τύποι τεχνολογιών και διάφορα μοντέλα εμπλέκονται σε τεχνικές ανίχνευσης, όπως μοντέλα RGB και HSI για την ανίχνευση καπνού και φωτιάς.

Η μελέτη [69] αναζητά μια εναλλακτική λύση, αναπτύσσοντας ένα σύστημα αναγνώρισης πυρκαγιάς που βασίζεται στην όραση. Το σύστημα λειτουργεί σε περιβάλλον MATLAB και βασίζεται σε ένα σύνολο κανόνων που αναπτύχθηκαν για τον προσδιορισμό της τιμής των στοιχείων R, G, B, Y, Cb και Cr σε μια εικόνα. Ο χρωματικός χώρος YCbCr δείχνει την καλύτερη απόδοση σε σύγκριση με το RGB, επειδή μπορεί να διαχωρίσει τη φωτεινότητα από τον χρωματισμό πιο αποτελεσματικά από τον χρωματικό χώρο RGB. Στο [70] παρουσιάζεται ένα σύστημα βιντεοεπιτήρησης, από το οποίο εξάγονται ορισμένα χωροχρονικά χαρακτηριστικά (χρώμα και κίνηση). Δεν γίνεται δηλαδή μόνο ανίχνευση της ίδιας της φωτιάς αλλά μελετάται και η εσωτερική γεωγραφία της πυρκαγιάς και η χρονική της ανάπτυξη. Στην εργασία [71] παρουσιάζονται τεχνικές όρασης υπολογιστή όπου γίνεται μέτρηση των ιδιοτήτων δασικής πυρκαγιάς (μέτωπο πυρκαγιάς, ύψος φλόγας, γωνία κλίσης φλόγας, πλάτος βάσης πυρκαγιάς). Το σύστημα υπολογίζει ένα τρισδιάστατο μοντέλο αντίληψης της φωτιάς και θα μπορούσε επίσης να χρησιμοποιηθεί για την απεικόνιση της εξέλιξης της πυρκαγιάς σε απομακρυσμένα συστήματα υπολογιστών. Το παρουσιαζόμενο σύστημα ενσωματώνει την επεξεργασία εικόνων από οπτικές και υπέρυθρες κάμερες. Εφαρμόζει τεχνικές σύντηξης αισθητήρων που περιλαμβάνουν επίσης αισθητήρες τηλεμετρίας και GPS.

Στο άρθρο [72], προτείνεται ένας ανιχνευτής πυρκαγιάς σε πραγματικό χρόνο που συνδυάζει πληροφορίες αντικειμένων στο προσκήνιο με στατιστικά έγχρωμων εικονοστοιχείων πυρκαγιάς. Το απλό προσαρμοστικό μοντέλο φόντου της σκηνής δημιουργείται χρησιμοποιώντας τρεις γκαουσιανές κατανομές, όπου κάθε κατανομή αντιστοιχεί στα στατιστικά εικονοστοιχείων στο αντίστοιχο κανάλι χρώματος. Οι πληροφορίες προσκηνίου εξάγονται χρησιμοποιώντας προσαρμοστικό αλγόριθμο αφαίρεσης φόντου και στη συνέχεια επαληθεύονται από το στατιστικό χρωματικό μοντέλο πυρκαγιάς για να προσδιοριστεί εάν το αντικείμενο προσκηνίου που ανιχνεύτηκε είναι υποψήφιο για πυρκαγιά ή όχι. Ένα γενικό μοντέλο χρώματος φωτιάς κατασκευάζεται με στατιστική ανάλυση των δειγμάτων εικόνων που περιέχουν εικονοστοιχεία πυρκαγιάς. Ο ρυθμός επεξεργασίας καρέ του ανιχνευτή είναι περίπου 40 fps με μέγεθος εικόνας 176 × 144 εικονοστοιχεία, και ο σωστός ρυθμός ανίχνευσης του αλγορίθμου είναι 98,89%. Στην εργασία [73], προτείνεται μια μέθοδος ανίχνευσης και παρακολούθησης δασικών πυρκαγιών βασισμένη σε μη επανδρωμένο εναέριο όχημα UAV. Αναπτύσσεται ένα σύνολο αλγορίθμων ανίχνευσης και παρακολούθησης δασικών πυρκαγιών που περιλαμβάνει φιλτράρισμα διάμεσου, μετατροπή χρωματικού χώρου, τμηματοποίηση κατωφλίου Otsu, μορφολογικές λειτουργίες και μετρητή κηλίδων. Η βασική ιδέα της προτεινόμενης μεθόδου είναι η υιοθέτηση του καναλιού «a» στο χρωματικό μοντέλο Lab για την εξαγωγή εικονοστοιχείων πυρκαγιάς χρησιμοποιώντας τα χρωματικά χαρακτηριστικά της φωτιάς.

Η άλλη 'ομάδα' παραδοσιακών προσεγγίσεων σε αυτόν τον τομέα επικεντρώνεται στη προσέγγιση όπου συνδυάζονται χρωματικές και χρονικές πληροφορίες υψηλής τάξης. Η εργασία [74] παρουσιάζει ένα αυτόματο σύστημα ανίχνευσης πυρκαγιάς σε ακολουθίες βίντεο. Προηγούμενες μέθοδοι ανίχνευσης πυρκαγιάς, οι οποίες χρησιμοποιούν οπτικές πληρο-

φορίες υποφέρουν από την αδυναμία να αντιμετωπίσουν μια κινούμενη κάμερα ή μια κινούμενη σκηνή. Προτείνεται ένα σύστημα που χρησιμοποιεί πληροφορίες χρώματος και κίνησης, που υπολογίζονται από ακολουθίες βίντεο για τον εντοπισμό της φωτιάς. Αυτό γίνεται χρησιμοποιώντας πρώτα μια προσέγγιση που βασίζεται στη δημιουργία ενός χρωματικού ιστογράμματος εξομάλυνσης κατά Gauss για την ανίχνευση των πυρόχρωμων εικονοστοιχείων και, στη συνέχεια, χρησιμοποιώντας μια χρονική παραλλαγή των εικονοστοιχείων για να προσδιοριστεί ποια από αυτά τα εικονοστοιχεία είναι πραγματικά εικονοστοιχεία φωτιάς. Στη συνέχεια, ορισμένα πλαστά εικονοστοιχεία πυρκαγιάς αφαιρούνται αυτόματα χρησιμοποιώντας μια λειτουργία 'διάβρωσης' και ορισμένα εικονοστοιχεία πυρκαγιάς που λείπουν εντοπίζονται, χρησιμοποιώντας τη μέθοδο ανάπτυξης περιοχής.

Στην εργασία [75], παρουσιάζονται φασματικά, χωρικά και χρονικά μοντέλα περιοχών πυρκαγιάς σε εικόνες. Το φασματικό μοντέλο αναπαρίσταται ως προς την πυκνότητα χρωματικής πιθανότητας των εικονοστοιχείων πυρκαγιάς. Το χωρικό μοντέλο εντοπίζει τη δομή 'χωρικά', μέσα σε μια περιοχή πυρκαγιάς. Το σχήμα μιας περιοχής πυρκαγιάς αναπαρίσταται μέσω της χωρικής συχνότητας του περιγράμματος της περιοχής χρησιμοποιώντας τους συντελεστές Fourier. Οι χρονικές αλλαγές σε αυτούς τους συντελεστές χρησιμοποιούνται ως χρονικές αλλαγές της πυρκαγιάς. Πιο συγκεκριμένα, χρησιμοποιείται ένα μοντέλο αυτόματης παλινδρόμησης της σειράς συντελεστών Fourier. Το άρθρο [76] προτείνει μια νέα μέθοδο για την ανίχνευση φλόγας σε βίντεο με την επεξεργασία των δεδομένων που παράγονται από μια συνηθισμένη κάμερα που παρακολουθεί μια σκηνή. Εκτός από τις ενδείξεις κίνησης και χρώματος, ανιχνεύεται επίσης η φλόγα που τρεμοπαίζει, χρησιμοποιώντας ένα κρυφό μοντέλο Markov. Τα μοντέλα Markov χρησιμοποιούνται για τη διάκριση της φλόγας από άλλα κινούμενα αντικείμενα με χρώμα φλόγας. Οι χωρικές, χρωματικές παραλλαγές στη φλόγα αξιολογούνται επίσης από τα ίδια μοντέλα Markov. Αυτές οι ενδείξεις συνδυάζονται για να καταλήξουν σε μια τελική απόφαση.

Η εργασία [77] παρουσιάζει μια τεχνική επεξεργασίας εικόνας για αυτόματη ανίχνευση πυρκαγιάς σε πραγματικό χρόνο σε εικόνες βίντεο. Ο υποκείμενος αλγόριθμος βασίζεται στη χρονική διακύμανση της έντασης της φωτιάς που καταγράφεται από έναν οπτικό αισθητήρα εικόνας. Οι πλήρεις ακολουθίες εικόνων αναλύονται για να επιλεγεί μια υποψήφια περιοχή φλόγας. Εξάγονται χαρακτηριστικά πυρκαγιάς από την υποψήφια περιοχή και συνδυάζονται για να προσδιοριστεί η παρουσία ή απουσία πυρκαγιάς. Ο συναγερμός πυρκαγιάς ενεργοποιείται αν το μοτίβο πυρκαγιάς επιμένει για κάποιο χρονικό διάστημα. Η εργασία [78] προτείνει μια νέα μέθοδο για την ανίχνευση πυρκαγιάς/φλόγας σε πραγματικό χρόνο με την επεξεργασία των δεδομένων βίντεο που παράγονται από μια συνηθισμένη κάμερα που παρακολουθεί μια σκηνή. Εκτός από τις ενδείξεις κίνησης και χρώματος, ανιχνεύεται και εδώ το τρεμόπαιγμα της φλόγας με την ανάλυση του βίντεο στον τομέα κυματιδίων. Αν και το τρεμόπαιγμα της φλόγας γενικά δεν είναι ημιτονοειδές ή περιοδικό υπό όλες τις συνθήκες, έχει παρατηρηθεί μια συχνότητα 10 Hz σε κάποιες μελέτες παρατήρησης. Η σχεδόν-περιοδική συμπεριφορά της φλόγας ανιχνεύεται με την εκτέλεση χρονικού μετασχηματισμού κυματιδίων. Οι χρωματικές παραλλαγές στις περιοχές φλόγας ανιχνεύονται με τον υπολογισμό του χωρικού μετασχηματισμού κυματιδίων κινούμενων περιοχών με χρώμα φωτιάς. Μια άλλη ένδειξη που χρησιμοποιείται στον αλγόριθμο πυρανίχνευσης είναι η ανωμαλία του ορίου της περιοχής με το χρώμα της φωτιάς. Όλα τα παραπάνω στοιχεία συνδυάζονται για να προκύψει

μια τελική απόφαση.

Αυτές οι παραδοσιακές μέθοδοι ενώ δίνουν επαρκή απόδοση δεν υπερέρχουν από τεχνικές μηχανικής μάθησης γιατί δεν είναι απαλλαγμένες από σφάλματα και δεν είναι πλήρως αξιόπιστες. Για παράδειγμα, οι μέθοδοι σύγκρισης τιμών RGB, που συνήθως θεωρούν μια τιμή κατωφλίου για την ανίχνευση πυρκαγιάς, υπάρχει περίπτωση να ανιχνεύσουν το ηλιοβασίλεμα και την ανατολή του ηλίου ως ψευδώς θετικό αποτέλεσμα.

### 3.1.2 Σύγχρονες προσεγγίσεις

Οι αλγόριθμοι μηχανικής μάθησης έχουν χρησιμοποιηθεί πρόσφατα για την ανίχνευση πυρκαγιάς χρησιμοποιώντας ακόμη και εναέριες εικόνες. Οι συγγραφείς στο [79] ανέπτυξαν μια προσέγγιση που βασίζεται στο Support Vector Machine (SVM) για την επίτευξη ανίχνευσης πυρκαγιών σε πραγματικό χρόνο. Ο SVM, ως ένας κλασικός αλγόριθμος μηχανικής μάθησης, μπορεί να επιτύχει καλή ακρίβεια για την ανίχνευση πυρκαγιάς. Όμως, η μέση ταχύτητα είναι πολύ πιο μικρή σε περίπλοκες καταστάσεις και η ακρίβεια της μεθόδου είναι πολύ χαμηλότερη από τον αναμενόμενο ρυθμό για την ανίχνευση πυρκαγιών στον πραγματικό κόσμο. Η εργασία [80] προτείνει μια μέθοδο πυρανίχνευσης που βασίζεται σε αισθητήρα όρασης για ένα σύστημα έγκαιρης προειδοποίησης πυρκαγιάς. Δημιουργείται ένα μοντέλο προσωρινής πυρκαγιάς με συντελεστές κυματιδίων και εφαρμόζεται ως είσοδο σε έναν ταξινομητή μηχανών διανυσμάτων υποστήριξης δύο τάξεων SVM με ακτινική συνάρτηση βάσης (RBF). Στη συνέχεια, ο ταξινομητής SVM χρησιμοποιείται για την τελική επαλήθευση εικονοστοιχείων πυρκαγιάς. Η προσέγγιση αυτή είναι πιο ανθεκτική σε θόρυβο, σε σχέση με τις παραδοσιακές μεθόδους πυρανίχνευσης σε εικόνες.

Στο άρθρο [81], προτείνεται ένας αλγόριθμος ανίχνευσης δασικών πυρκαγιών σε πραγματικό χρόνο χρησιμοποιώντας τεχνητά νευρωνικά δίκτυα, βασισμένος σε δυναμικά χαρακτηριστικά των περιοχών πυρκαγιάς, τμηματοποιημένες από εικόνες βίντεο. Η περιοχή πυρκαγιάς λαμβάνεται από την εικόνα με τη βοήθεια τιμών κατωφλίου στον χρωματικό χώρο HSV. Η περιοχή, η στρογγυλότητα και το περίγραμμα υπολογίζονται για περιοχές πυρκαγιάς από κάθε 5 συνεχή πλαίσια. Η μέση και η μέση τετραγωνική απόκλιση χρησιμοποιούνται ως δυναμικά χαρακτηριστικά και λαμβάνονται ως είσοδοι του τεχνητού νευρωνικού δικτύου. Το εκπαιδευμένο δίκτυο της BP μπορεί να βοηθήσει στον εντοπισμό της δασικής πυρκαγιάς, ακόμη και να τη διακρίνει από κινούμενο αυτοκίνητο ή σημαία με κόκκινο χρώμα. Στην [82] διερευνάται η χρήση μιας μη χρονικής προσέγγισης με τη συνδυασμένη χρήση περιγραφέντων χαρακτηριστικών χρώματος-υφής ως είσοδο σε ένα εκπαιδευμένο ταξινομητή που είναι ή ένα δέντρο αποφάσεων ή ρηχά νευρωνικά δίκτυα. Η προτεινόμενη προσέγγιση είναι μια ισχυρή λύση σε πραγματικό χρόνο σε αυτό το πεδίο, αφού επιτυγχάνει περίπου 80 – 90% μέση πραγματική θετική ανίχνευση και γύρω στο 7 – 8% ψευδώς θετική.

Στο [83] παρουσιάζεται ένας αλγόριθμος ανίχνευσης πυρκαγιάς που βασίζεται σε πλαίσια με διαδικτυακή μάθηση ακραίων στοιχείων. Στον προτεινόμενο αλγόριθμο, οι υποψήφιος φωτιές λαμβάνονται με τη μορφή πλαισίων, επειδή έχουν πιο διακριτικό σχήμα από ολόκληρη τη φωτιά. Έτσι, ο ταξινομητής σχήματος μπορεί να αναγνωρίσει σωστά τις υποψήφιος φωτιές, από ακραία σημεία που μοιάζουν με φωτιά. Επιπλέον, προτείνεται ένα διαδικτυακό σχήμα εκμάθησης ακραίων στοιχείων που χειρίζεται την ανωμαλία της πυρκαγιάς με βάση

την επαναληψιμότητα του σχήματος στο χρόνο. Η εργασία [84] παρουσιάζει μια συγκριτική ανάλυση για το πρόβλημα της ανίχνευσης πυρκαγιάς μεταξύ μοντέλων βασισμένων σε χρώμα για την εξαγωγή κανόνων και ενός αλγορίθμου μηχανικής μάθησης. Η τεχνική μηχανικής μάθησης που χρησιμοποιείται εδώ είναι η λογιστική παλινδρόμηση είναι μία από τις απλούστερες τεχνικές στη μηχανική μάθηση. Η προτεινόμενη τεχνική μηχανικής μάθησης ξεπερνά τους υπάρχοντες αλγόριθμους αιχμής στη βιβλιογραφία, με βάση το χρώμα, σε όλα σχεδόν τα σενάρια.

### 3.1.3 Προσεγγίσεις βαθιάς μάθησης

Τα νευρωνικά δίκτυα έχουν χρησιμοποιηθεί με επιτυχία σε πολλούς διαφορετικούς τομείς όπως η επεξεργασία φυσικής γλώσσας, η αναγνώριση ομιλίας, η ανάλυση κειμένου και ιδιαίτερα η ταξινόμηση εικόνων. Προηγουμένως, είδαμε να χρησιμοποιούνται πολλοί και διαφορετικοί αλγόριθμοι, ο ένας πάνω στον άλλον για να εξαχθούν τα σχετικά χαρακτηριστικά. Αντί αυτής της τεχνικής, μπορεί να χρησιμοποιηθεί ένα δίκτυο που μαθαίνει τα χαρακτηριστικά αυτά από μόνο του. Συμπεραίνουμε λοιπόν, ένα πρόβλημα ανίχνευσης πυρκαγιάς είναι κατάλληλο για βαθιά μάθηση. Ενώ υπάρχουν πολλές σημαντικές συνεισφορές από ερευνητές για την ανάπτυξη ενός συστήματος που να ανιχνεύει με ακρίβεια τη φωτιά στο περιβάλλον, τα πιο επιτυχημένα συστήματα περιλαμβάνουν βαθιά συνελκτικά δίκτυα (DCNN).

Μια τέτοια προσέγγιση ακολουθείται στο [85], όπου σε αυτό το πρόβλημα χρησιμοποιείται ένας αλγόριθμος εκμάθησης για να εξαχθούν τα χρήσιμα χαρακτηριστικά αντί να χρησιμοποιηθεί ένας ειδικός για τη δημιουργία τους. Προτείνεται ένα CNN ως ανιχνευτής πλαισίου πυρκαγιάς/καπνού. Αυτό το δίκτυο λειτουργεί απευθείας σε ακατέργαστο πλαίσιο RGB και εκτελεί εξαγωγή και ταξινόμηση χαρακτηριστικών εντός της ίδιας αρχιτεκτονικής. Η αρχιτεκτονική ταξινόμησης είναι κλασική για το συνελκτικό νευρωνικό δίκτυο, συνδυάζοντας συνέλιξη και υποδειγματοληψία μεγίστου. Η εργασία [86] προτείνει έναν αλγόριθμο ανίχνευσης καπνού με βάση τα χαρακτηριστικά κίνησης του καπνού και τα CNN. Αρχικά, προτείνεται ένας αλγόριθμος ανίχνευσης κινούμενου αντικειμένου που βασίζεται στη δυναμική ενημέρωση του παρασκηνίου και στο σκοτεινό κανάλι για τον εντοπισμό των ύποπτων περιοχών καπνού. Στη συνέχεια, τα χαρακτηριστικά της ύποπτης περιοχής εξαγονται αυτόματα από το CNN, στο οποίο πραγματοποιείται η αναγνώριση καπνού. Για το πρόβλημα ότι η περιοχή καπνού είναι σχετικά μικρή στο πρώιμο στάδιο της δημιουργίας καπνού, προτείνεται η στρατηγική της σιωπηρής μεγέθυνσης των ύποπτων περιοχών, η οποία βελτιώνει την έγκαιρη ανίχνευση καπνού.

Στην εργασία [87] προτείνεται μια μέθοδος βαθιάς μάθησης για την ανίχνευση δασικών πυρκαγιών. Πρώτα εκπαιδεύεται ένας ταξινομητής πυρκαγιάς πλήρους εικόνας και στη συνέχεια εφαρμόζεται ένας ταξινομητής πλαισίου εάν η εικόνα ταξινομηθεί ότι περιέχει φωτιά (cascade CNN fire detector). Στην εργασία [88] προτείνεται η χρήση βαθύτερων συνελκτικών νευρωνικών δικτύων για την ανίχνευση πυρκαγιάς σε εικόνες. Χρησιμοποιούνται δύο state of the art προεκπαιδευμένα βαθιά CNN, VGG16 και Resnet50, για να αναπτυχθεί το σύστημα πυρανίχνευσης. Χρησιμοποιείται το VGG16 έναντι του VGG19 γιατί κάνει λιγότερη ώρα να εκπαιδευτεί και το σύνολο δεδομένων δεν είναι τόσο περίπλοκο όσο το ImageNet. Τα μοντέλα

αυτά ελέγχονται σε ένα μη ισορροπημένο σύνολο δεδομένων, με εικόνες που είναι δύσκολο να ταξινομηθούν, οι οποίες έχουν συγκεντρωθεί για να αναπαραχθούν σενάρια πραγματικού κόσμου. Με την προσθήκη πλήρως συνδεδεμένων επιπέδων αυξάνεται η ακρίβεια της πυρανίχνευσης, αλλά αυξάνεται επίσης και ο χρόνος εκπαίδευσης. Το Resnet50 αποδείχτηκε να έχει ελαφρώς καλύτερη απόδοση από το VGG16. Ακόμη δοκιμάστηκαν εναλλακτικές, απλοποιημένες υλοποιήσεις του VGG16 με 4 fully connected layers και του ResNet.

Στην εργασία [89] διερευνάται η αυτόματη ανίχνευση εικονοστοιχείων πυρκαγιάς, σε εικόνες βίντεο (ή στατικές) εντός ορίων πραγματικού χρόνου, χωρίς να βασίζεται σε πληροφορίες χρονικής σκηνής. Δοκιμάζοντας τις τρεις αρχιτεκτονικές CNN (AlexNet, VGG-13, InceptionV1) στην εργασία ανίχνευσης πυρκαγιάς, πραγματοποιείται μια πειραματική αξιολόγηση των οριακά καλύτερων σε απόδοση αρχιτεκτονικών AlexNet και InceptionV1. Προτείνονται μειωμένες αρχιτεκτονικές οι οποίες προσφέρουν επιπλέον 3-4 φορές αύξηση στην υπολογιστική απόδοση, προσφέροντας επεξεργασία έως και 17 fps σε σύγχρονο υλικό, ανεξάρτητα από χρονικές πληροφορίες. Προτάθηκαν 6 διαφορετικές εκδοχές του AlexNet και 8 εκδοχές του InceptionV1 αφαιρώντας στην καθεμιά έναν αριθμό επιπέδων. Οι καλύτερες αρχιτεκτονικές σε απόδοση που προέκυψαν είναι το FireNet, το οποίο βασίζεται στο AlexNet, και το InceptionV1-OnFire, το οποίο βασίζεται σε μια μειωμένη αρχιτεκτονική InceptionV1.

Στην εργασία [90] εξερευνούνται διαφορετικές αρχιτεκτονικές CNN και παραλλαγές τους για δύο προβλήματα ανίχνευσης πυρκαγιάς. Το πρώτο είναι η μη χρονική δυαδική ανίχνευση πυρκαγιάς για τον προσδιορισμό της ύπαρξης πυρκαγιάς σε ένα συγκεκριμένο πλαίσιο και το δεύτερο ο εντοπισμός superpixel εντός του πλαισίου για τον προσδιορισμό της ακριβούς θέσης της πυρκαγιάς μέσα σε αυτό το πλαίσιο. Δοκιμάζονται αρχιτεκτονικές CNN με μειωμένη πολυπλοκότητα όπως Inception, ResNet, EfficientNet. Ορίζονται παραλλαγές αυτών όπως τα InceptionV2, InceptionV3, InceptionV4, Inception-ResNet καθώς και παραλλαγές της υβριδικής αρχιτεκτονικής v1, v2 των οποίων η μόνη διαφορά έγκειται στις ρυθμίσεις υπερπαραμέτρων. Το Inception-ResNet έχει αποδειχθεί ότι επιτυγχάνει ανώτερη απόδοση με μικρότερο αριθμό εποχών. Με εξαντλητικούς πειραματισμούς, για το πλήρες σύνολο των παραλλαγών που περιγράφονται, προτείνονται οι ακόλουθες δύο αρχιτεκτονικές, με μέγιστη μειωμένη πολυπλοκότητα: α) το InceptionV3-OnFire δίκτυο, εμπνευσμένο από την απόδοση της παραλλαγής InceptionV3V09 και β) η InceptionV4-OnFire έκδοση που είναι τριών επιπέδων, εμπνευσμένο από το InceptionV4 που βασίζεται στην παραλλαγή InceptionV4v05.

Στο άρθρο [91] προτείνεται ένας νέος αλγόριθμος CNN για την επίτευξη ανίχνευσης πυρκαγιάς και καπνού σε εικόνα, με υψηλή ακρίβεια. Για να λυθεί το πρόβλημα υπερπροσαρμογής, που προκαλείται από την εκπαίδευση του δικτύου σε ένα περιορισμένο σύνολο δεδομένων, βελτιώνεται το πλήθος των δεδομένων με τη χρήση GAN. Το CNN που προτάθηκε χρησιμοποιεί ιδέες από το VGGNet. Αποτελείται από 12 επίπεδα, 6 συνελκτικά, 3 επίπεδα συγκέντρωσης για εξαγωγή χαρακτηριστικών, και 3 πλήρως συνδεδεμένα στρώματα, στο τέλος, για ταξινόμηση. Η κύρια τροποποίηση που έγινε στο δίκτυο, είναι η χρήση της συνάρτησης προσαρμοστικής, τμηματικής, γραμμικής ενεργοποίησης (APL), αντί της χρήσης των παραδοσιακών ReLU σε συνελκτικά επίπεδα. Αυτό βελτίωσε την απόδοση του δικτύου, αλλά αυξήθηκε ο χρόνος εκπαίδευσης, καθώς διαθέτει ορισμένες πρόσθετες λειτουργίες. Μια άλλη αλλαγή είναι η χρήση διαφορετικού αριθμού φίλτρων, μικρού μεγέθους 3x3 για όλα τα συνελκτικά επίπεδα.

Στην εργασία [92], διερευνώνται διαφορετικές αρχιτεκτονικές CNN και παραλλαγές τους, για τη μη χρονική ανίχνευση σε πραγματικό χρόνο, των περιοχών εικονοστοιχείων πυρκαγιάς σε εικόνες βίντεο (ή στατικές). Προτείνονται δύο αρχιτεκτονικές CNN μειωμένης πολυπλοκότητας NasNet-A-OnFire και ShuffleNetV2-OnFire μέσω πειραματικής ανάλυσης για τη βελτιστοποίηση της υπολογιστικής απόδοσης. Επιτυγχάνουν ακρίβεια 95% για δυαδική ταξινόμηση (fire, no-fire) πλήρους πλαισίου και 97% για εντοπισμό superpixel, με χρόνο εκτέλεσης 40 fps και 18 fps αντίστοιχα.

Στο άρθρο [93] προτείνεται η χρήση μιας προεκπαιδευμένης αρχιτεκτονικής MobileNetV2 για την υλοποίηση της εκμάθησης μέσω μεταφοράς, η οποία απαιτεί μικρότερο σύνολο δεδομένων και μειώνει την υπολογιστική πολυπλοκότητα, ενώ δεν διακυβεύει την ακρίβεια. Επιπλέον, πραγματοποιείται ένα αποτελεσματικό pipeline αύξησης δεδομένων για την προσομοίωση ορισμένων ακραίων σεναρίων, τα οποία θα μπορούσαν να κάνουν την προσέγγιση πιο εύρωστη. Τα αποτελέσματα των δοκιμών καταδεικνύουν ότι η μέθοδος αυτή διατηρεί υψηλή ακρίβεια αναγνώρισης σε διαφορετικές καταστάσεις.

Το άρθρο [94] παρέχει ένα σύνολο δεδομένων εικόνων πυρκαγιάς που συλλέχθηκε από drones κατά τη διάρκεια μιας προδιαγεγραμμένης καύσης υπολειμμάτων σε ένα πευκοδάσος της Αριζόνα. Η εργασία προτείνει λύσεις σε δύο προβλήματα μηχανικής μάθησης: α) Δυαδική ταξινόμηση πλαισίων βίντεο με βάση την παρουσία (και απουσία) φλόγας φωτιάς. Αναπτύχθηκε μια μέθοδος Τεχνητού Νευρωνικού Δικτύου (ANN) που πέτυχε ακρίβεια ταξινόμησης 76%. Αυτή βασίζεται σε έναν αλγόριθμο DCNN, γνωστό ως Xception που προτάθηκε από τη Google-Keras.

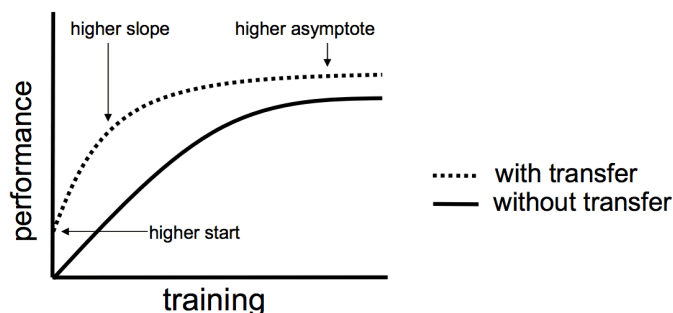
## 3.2 Μεταφορά Μάθησης

*Μεταφορά μάθησης* (Transfer Learning) ονομάζεται το πρόβλημα έρευνας της μηχανικής μάθησης, το οποίο βασίζεται στην αξιοποίηση της γνώσης, που αποκτήθηκε από την εκπαίδευση ενός συστήματος, σε ένα διαφορετικό αλλά σχετικό πρόβλημα. Για παράδειγμα, η γνώση από ένα σύστημα που αναγνωρίζει αεροπλάνα μπορεί να χρησιμοποιηθεί σε ένα άλλο σύστημα που αναγνωρίζει ελικόπτερα.

Η μάθηση με μεταφορά είναι μια βελτιστοποίηση και μια συντόμευση για εξοικονόμηση χρόνου ή καλύτερη απόδοση. Γενικά, δεν είναι προφανές ότι θα υπάρχει όφελος από τη χρήση της μάθησης μέσω μεταφοράς παρά μόνο αφού αναπτυχθεί και αξιολογηθεί το μοντέλο. Αλλά στις περισσότερες περιπτώσεις, δίνει καλύτερα αποτελέσματα από ένα μοντέλο που έχει εκπαιδευτεί από την αρχή (Εικόνα 3.1).

Τα βαθιά νευρωνικά δίκτυα χρησιμοποιούνται για να επιλύσουν περίπλοκα προβλήματα, τα οποία απαιτούν μεγάλο πλήθος δεδομένων για να εκπαιδευτούν. Συχνά όμως, αντιμετωπίζουμε χρονικούς και υπολογιστικούς περιορισμούς και η εκπαίδευση ενός τέτοιου συστήματος είναι ανέφικτο να γίνει απ΄την αρχή, σε ένα μηχάνημα περιορισμένων δυνατοτήτων. Η μάθηση μέσω μεταφοράς είναι αυτή που μπορεί να λύσει αυτό το πρόβλημα.

Αφορμώμενοι από τις δύο τελευταίες σχετικές εργασίες της προηγούμενης ενότητας, χρησιμοποιούμε στο πλαίσιο της παρούσας εργασίας, τα μοντέλα Xception και MobileNet2. Παρακάτω, παρουσιάζονται αναλυτικά ως προς τη δομή και λειτουργία τους και θα χρησιμοποιηθούν για την εκπαίδευση, με transfer learning, του συνόλου δεδομένων στο πρόβλημα



Εικόνα 3.1: Αντιπαράθεση της μάθησης μέσω μεταφοράς και της εκπαίδευσης από την αρχή (Πηγή [15])

της ταξινόμησης για ύπαρξη ή όχι φωτιάς.

### 3.2.1 Xception Model

Το Xception μοντέλο υλοποιήθηκε από ερευνητές της Google και παρουσιάστηκε πρώτη φορά το 2016 από τον Francois Chollet [95]. Είναι μια αρχιτεκτονική βαθιών συνελκτικών δικτύων εμπνευσμένη από το μοντέλο Inception και γι'αυτό ονομάστηκε έτσι ως “Extreme Inception”. Η διαφορά είναι ότι τα στοιχεία του Inception έχουν αντικατασταθεί από κατά βάθος διαχωρίσιμες συνελίξεις.

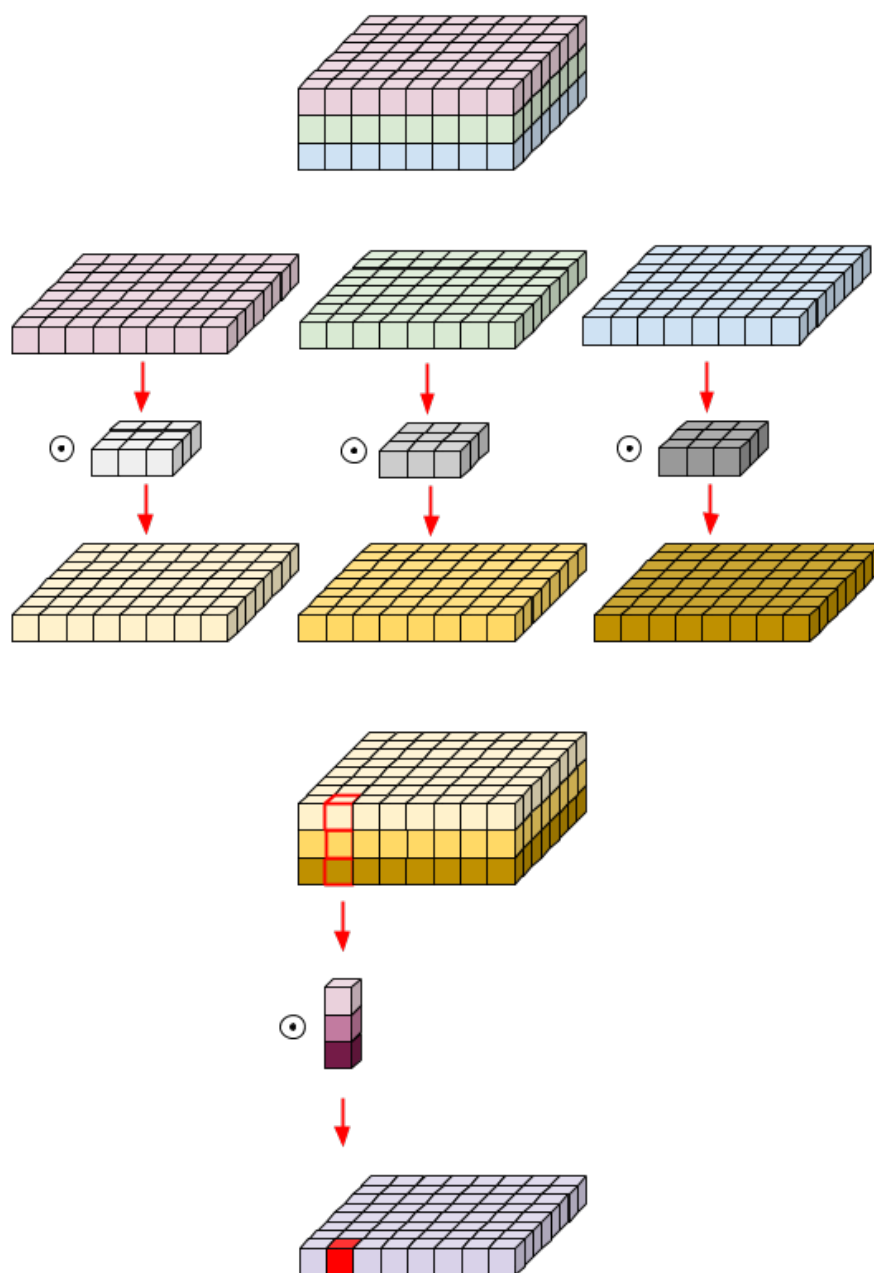
Για να γίνει κατανοητή η αρχιτεκτονική του Xception θα εξηγηθεί αρχικά η διαχωρίσιμη συνέλιξη και η διαφορά της από την απλή συνέλιξη. Η διαχωρίσιμη ανά κανάλι συνέλιξη μπορεί να είναι δύο ειδών, χωρική και κατά βάθος.

Η *χωρική διαχωρίσιμη συνέλιξη*, λόγω των περιορισμών της, δεν χρησιμοποιείται ευρέως στη βαθιά μάθηση. Ονομάστηκε έτσι επειδή ασχολείται με τις χωρικές διαστάσεις της εικόνας και του πυρήνα, το πλάτος και το ύψος. Μια χωρική διαχωρίσιμη συνέλιξη χωρίζει έναν πυρήνα σε δύο, μικρότερους πυρήνες. Ένα συνηθισμένο παράδειγμα είναι αυτό του διαχωρισμού ενός πυρήνα  $3 \times 3$  σε πυρήνα  $3 \times 1$  και  $1 \times 3$ . Ο κύριος περιορισμός αυτής της συνελίξης είναι ότι δεν διαχωρίζονται όλοι οι πυρήνες στα δύο.

Το ζήτημα αυτό το ‘επιλύει’ η *κατά βάθος διαχωρίσιμη συνέλιξη*. Ονομάστηκε έτσι επειδή ασχολείται με τις χωρικές διαστάσεις και επιπλέον με τη διάσταση βάθους, δηλαδή τον αριθμό των καναλιών. Παρόμοια με τη χωρική συνέλιξη, η κατά βάθος διαχωρίσιμη συνέλιξη χωρίζει έναν πυρήνα σε 2 ξεχωριστούς πυρήνες που κάνουν δύο συνελίξεις, τη συνέλιξη κατά βάθος και τη σημειακή συνέλιξη. Στην αρχή, γίνεται η συνέλιξη κατά βάθος (Depthwise Convolution), κατά την οποία δίνουμε στην εικόνα εισόδου μια συνέλιξη χωρίς να αλλάξουμε το βάθος, χρησιμοποιώντας 3 πυρήνες, όσα και τα κανάλια εισόδου. Κάθε πυρήνας επαναλαμβάνει ένα κανάλι της εικόνας, λαμβάνοντας τα βαθμωτά γινόμενα και η στοιβάξη αυτών των εικόνων μαζί δημιουργεί την εικόνα εξόδου. Στη συνέχεια, έπεται η σημειακή συνέλιξη (Pointwise Convolution), στην οποία χρησιμοποιείται ένας πυρήνας  $1 \times 1$  που επαναλαμβάνεται σε κάθε σημείο. Αυτός ο πυρήνας έχει βάθος όσα τα κανάλια της εικόνας εισόδου. Για παράδειγμα, επαναλαμβάνουμε έναν πυρήνα  $1 \times 1 \times 3$  οκτώ φορές στην εικόνα  $8 \times 8 \times 3$ , για να λάβουμε μια εικόνα  $8 \times 8 \times 1$ . Στην Εικόνα 3.2 περιγράφεται σχηματικά η όλη διαδικασία.

Το κύριο πλεονέκτημα της κατά βάθος διαχωρίσιμης ανα κανάλι συνελίξης είναι ότι





**Εικόνα 3.2:** Παράδειγμα της κατά βάθος διαχωρίσιμης συνέλιξης, όπου ο χάρτης εισόδου έχει διαστάσεις  $8 \times 8 \times 3$ , ο πυρήνας έχει διαστάσεις  $3 \times 3 \times 3$  και ο χάρτης εξόδου  $8 \times 8 \times 1$  (Πηγή [16]).

με τη χρήση της, μειώνεται δραστικά ο αριθμός των πολλαπλασιασμών άρα και των παραμέτρων. Στην κανονική συνέλιξη, η εικόνα μετασχηματίζεται τόσες φορές όσα τα κανάλια που θέλουμε. Στη διαχωρίσιμη συνέλιξη, η εικόνα μετασχηματίζεται μόνο μία φορά, στη συνέλιξη σε βάθος. Στη συνέχεια, η μετασχηματισμένη εικόνα επιμηκύνεται όσες φορές είναι τα επιθυμητά κανάλια. Έτσι, εξοικονομείται υπολογιστική ισχύς.

Πιο συγκεκριμένα, θα προσπαθήσουμε να εξηγήσουμε το υπολογιστικό κόστος της κάθε περίπτωσης. Γι' αυτό το σκοπό, ας θεωρήσουμε ένα επίπεδο συνέλιξης με είσοδο ένα χάρτη

χαρακτηριστικών F διαστάσεων  $H_F \times W_F \times M$ , όπου τα  $H_F, W_F$  αντιστοιχούν στο ύψος και το πλάτος των δεδομένων εισόδου και το M αντιστοιχεί στο πλήθος των καναλιών εισόδου (βάθος). Θεωρούμε επίσης ότι ο παραγόμενος χάρτης χαρακτηριστικών G έχει διαστάσεις  $H_G \times W_G \times N$  και ο πυρήνας της συνέλιξης K έχει διαστάσεις  $H_K \times W_K \times M \times N$ .

Σε μια εκτέλεση τυπικής συνέλιξης, ο χάρτης εξόδου υπολογίζεται όπως φαίνεται παρακάτω από την Εξίσωση 3.1:

$$G_{k,l,m} = \sum_{i,j,m} K_{i,j,m,n} F_{k+i-1,l+j-1,m} \quad (3.1)$$

,με υπολογιστικό κόστος που φαίνεται στην Εξίσωση 3.2:

$$H_K \cdot W_K \cdot M \cdot N \cdot H_F \cdot W_F \quad (3.2)$$

Αντίθετα, σε μια εκτέλεση διαχωρίσιμης ανά κανάλι συνέλιξης, στο πρώτο στάδιο η παραπάνω συνέλιξη επιμερίζεται σε M το πλήθος πυρήνες δύο διαστάσεων, με τον ενδιάμεσο χάρτη εξόδου G' για το m-στό κανάλι να δίνεται από την Εξίσωση 3.3:

$$G'_{k,l,m} = \sum_{i,j} K_{i,j,m} F_{k+i-1,l+j-1,m} \quad (3.3)$$

και συνολικό υπολογιστικό κόστος να είναι όπως φαίνεται παρακάτω στην Εξίσωση 3.4:

$$H_K \cdot W_K \cdot M \cdot H_F \cdot W_F \quad (3.4)$$

Στο δεύτερο βήμα, εκτελείται μια  $1 \times 1 \times M \times N$  συνέλιξη σε κάθε ένα από τα  $H_F \cdot W_F$  σημεία του ενδιάμεσου χάρτη εξόδου, με συνολικό υπολογιστικό κόστος που φαίνεται από την Εξίσωση 3.5:

$$M \cdot N \cdot H_F \cdot W_F \quad (3.5)$$

Για το σύνολο της εκτέλεσης της διαχωρίσιμης ανά κανάλι συνέλιξης, έχουμε υπολογιστικό κόστος που φαίνεται παρακάτω στην Εξίσωση 3.6:

$$H_K \cdot W_K \cdot M \cdot H_F \cdot W_F + M \cdot N \cdot H_F \cdot W_F \quad (3.6)$$

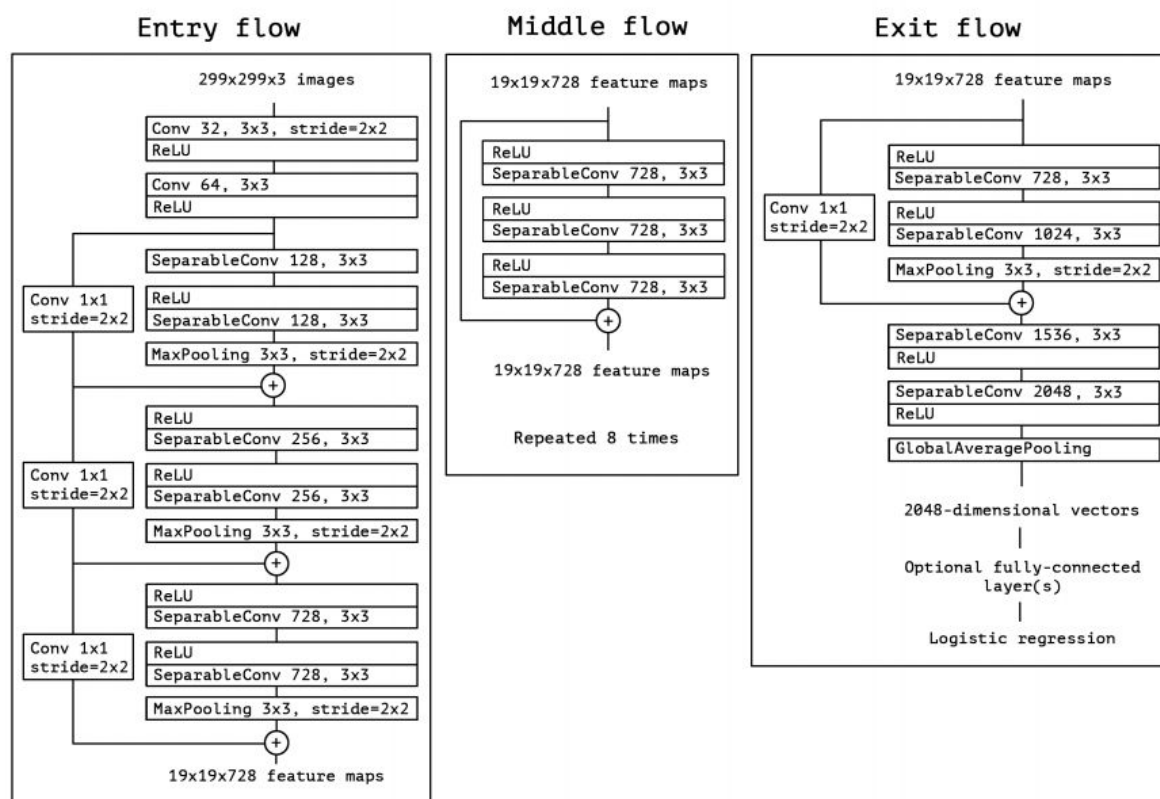
Ο συντελεστής μείωσης του υπολογιστικού κόστους υπολογίζεται από την Εξίσωση 3.7:

$$\frac{H_K \cdot W_K \cdot M \cdot H_F \cdot W_F + M \cdot N \cdot H_F \cdot W_F}{H_K \cdot W_K \cdot M \cdot N \cdot H_F \cdot W_F} = \frac{1}{N} + \frac{1}{H_K \cdot W_K} \quad (3.7)$$

και μπορεί να γίνει κατανοητό ότι για σχετικά μικρό βάθος εξόδου και μικρό μέγεθος φίλτρου, όπως  $N = 20$ ,  $H_K = 4$ ,  $W_K = 5$ , είναι εφικτή μια μείωση του υπολογιστικού κόστους κατά μια τάξη μεγέθους.

Η αρχιτεκτονική Xception αποτελείται συνολικά από 36 συνεκτικά επίπεδα. Αυτή η αρχιτεκτονική είναι μια γραμμική στοίβα συνελίξεων, κατά βάθος διαχωρίσιμων, αποτελείται από 14 μονάδες που συνδέονται με “residual” συνδέσεις. Σημειώνεται ότι μετά από κάθε επίπεδο κανονικής συνέλιξης και συνέλιξης διαχωρίσιμης κατά βάθος, ακολουθεί Batch

Normalization και ότι χρησιμοποιείται παντού στο δίκτυο, εκτός από την έξοδο, συνάρτηση ενεργοποίησης ReLU (Εικόνα 3.3).



Εικόνα 3.3: Αρχιτεκτονική Xception δικτύου (Πηγή [17])

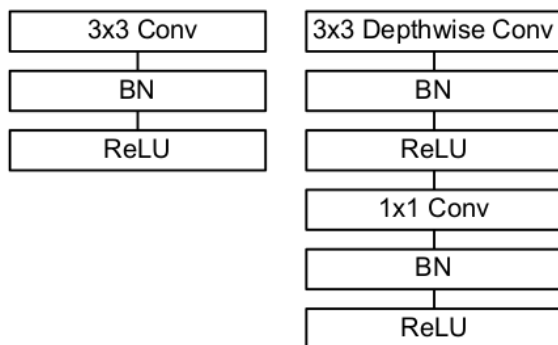
Η αρχιτεκτονική Xception, ξεπερνά ελαφρώς το Inception V3 στο σύνολο δεδομένων ImageNet και ξεπερνά σημαντικά το Inception V3 σε ένα μεγαλύτερο σύνολο δεδομένων ταξινόμησης εικόνων που περιλαμβάνει 350 εκατομμύρια εικόνες και 17.000 κλάσεις. Έχει τον ίδιο αριθμό παραμέτρων με το Inception V3 και τα κέρδη απόδοσης οφείλονται στην αποτελεσματικότερη χρήση αυτών.

### 3.2.2 MobileNetV2 Model

Το MobileNetV2 μοντέλο υλοποιήθηκε από ερευνητές της Google και παρουσιάστηκε πρώτη φορά το 2018. Στην προηγούμενη έκδοση, MobileNetV1, εισήχθηκε κατά βάθος διαχωρίσιμη συνέλιξη που μείωσε αποτελεσματικά τη πολυπλοκότητα και το μέγεθος του δικτύου. Το δίκτυο αυτό είναι κατάλληλο για φορητές συσκευές ή οποιεσδήποτε συσκευές με χαμηλή υπολογιστική ισχύ. Από την άλλη, το MobileNetV2 βασίζεται σε μια ανεστραμμένη δομή όπου οι συνδέσεις συντόμευσης βρίσκονται μεταξύ των λεπτών “bottleneck layers”. Το MobileNetV2 περιλαμβάνει αρχικό επίπεδο ‘συμφόρησης’ με 32 φίλτρα, ακολουθούμενο από 19 residual bottleneck layers. Παρακάτω θα αναλυθούν αναλυτικότερα τα δίκτυα αυτά και τα επιμέρους συστατικά τους.

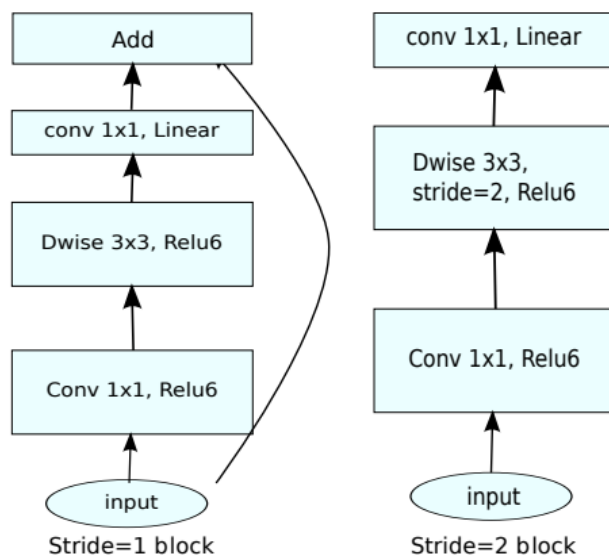
Το MobileNetV1 βασίζεται σε δύο βασικά δομικά συστατικά. Το πρώτο είναι το επίπεδο συνέλιξης το οποίο απαρτίζεται από εφαρμογή πυρήνα διαστάσεων  $3 \times 3$  και βάθους

αντίστοιχου του βάθους εισόδου. Ακολουθείται από ένα επίπεδο κανονικοποίησης και στη συνέχεια τροφοδοτεί ένα επίπεδο ενεργοποίησης ανορθωμένης γραμμικής μονάδας. Το δεύτερο είναι το επίπεδο διαχωρίσιμης συνέλιξης ανά κανάλι. Υλοποιείται από την προσθήκη ενός επιπέδου κανονικοποίησης και ενός επιπέδου ενεργοποίησης ανορθωμένης γραμμικής μονάδας, μετά από κάθε στάδιο μιας διαχωρίσιμης ανά κανάλι συνέλιξης με πυρήνα διαστάσεων  $3 \times 3$  (Εικόνα 3.4) [96].



Εικόνα 3.4: Δομικά στοιχεία αρχιτεκτονικής MobileNetV1 (Πηγή [18])

Όπως αναφέρθηκε και προηγουμένως, το MobileNetV2 βασίζει την υλοποίησή του στο δομικό σχήμα του μπλοκ συμφόρησης με κατάλοιπο (bottleneck residual block). Περιλαμβάνει ένα επίπεδο επέκτασης, που χρησιμοποιεί συνέλιξη ανά σημείο για να επεκτείνει τα χαρακτηριστικά εισόδου σε μεγαλύτερο βάθος ανάλυσης, μια διαχωρίσιμη ανά κανάλι συνέλιξη και ένα επίπεδο γραμμικής συμφόρησης. Κάθε επιμέρους επίπεδο ακολουθείται από ένα επίπεδο κανονικοποίησης και, πλην του επιπέδου γραμμικής συμφόρησης, από ένα επίπεδο ενεργοποίησης με ανορθωμένη γραμμική μονάδα (Εικόνα 3.5).



Εικόνα 3.5: Απεικόνιση μπλοκ συμφόρησης με κατάλοιπο. Παράδειγμα block με  $n > 1$  και  $s = 2$ . Στα δεξιά απεικονίζεται η δομή του πρώτου block και στα αριστερά η δομή όλων των μετέπειτα blocks (Πηγή [19]).

Το *επίπεδο συμφόρησης* (Bottleneck Layer) αποτελεί μια ειδική περίπτωση συνέλιξης ανά σημείο, όπου το βάθος του χάρτη εξόδου είναι μικρότερο από το βάθος της εισόδου. Με αυτό τον τρόπο, μπορεί να επιτευχθεί μείωση του βάθους της εξόδου και μείωση του πλήθους των παραμέτρων του συστήματος, αφού η έξοδος μειωμένου βάθους χρησιμοποιείται ως είσοδος στο επόμενο επίπεδο.

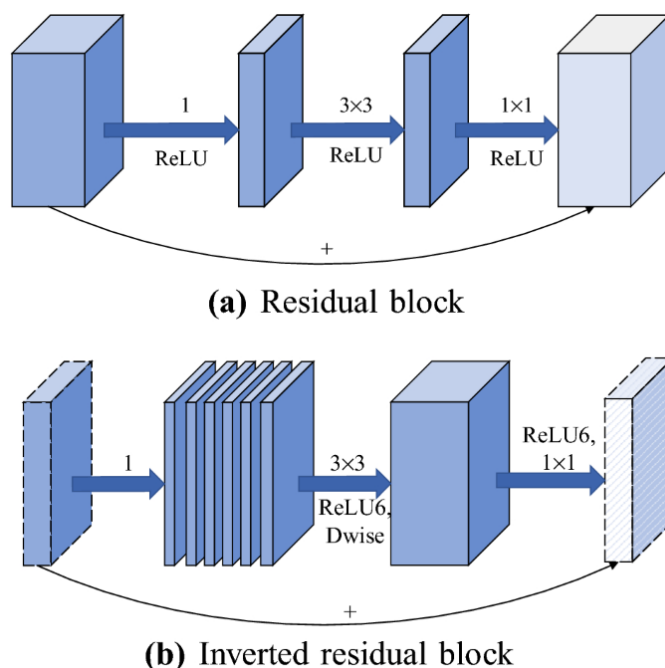
Το *επίπεδο γραμμικής συμφόρησης* (Linear Bottleneck Layer) αποτελεί ειδική περίπτωση του επιπέδου συμφόρησης, με την απαίτηση ότι δεν έχει εφαρμοστεί κάποιος μη-γραμμικός μετασχηματισμός στην είσοδό της, αφού η εφαρμογή μη-γραμμικών μετασχηματισμών αλλοιώνει το περιεχόμενο της πληροφορίας. Υπό περιπτώσεις, είναι εφικτή η διατήρηση της πληροφορίας μετά την εφαρμογή της ανορθωμένης γραμμικής μονάδας στην είσοδο, εφόσον υπάρχει επαρκές πλήθος μη-μηδενικών ενεργοποιήσεων και το μέγεθος της χρήσιμης πληροφορίας είναι αρκετά μικρό σε σχέση με το μέγεθος της εισόδου.

Η επίδραση που έχει η προσθήκη επιπέδων σε ένα δίκτυο, δεν είναι δεδομένο ότι θα έχει καλύτερη επίδοση. Για να αποφευχθεί η πιθανότητα χειροτέρευσης ενός πιο σύνθετου δικτύου, θα πρέπει να εξασφαλίσουμε ότι κάθε επέκταση του δικτύου είναι ικανή να αναπαραστήσει την εκφραστικότητα του αρχικού δικτύου. Η παραπάνω συμπεριφορά μπορεί να επιτευχθεί με την προσθήκη μιας σύνδεσης καταλοίπου (residual connection) μεταξύ δύο επιπέδων του δικτύου, με το περιεχόμενο της αρχής της σύνδεσης να προστίθεται στην έξοδο του τέλους της. Απαραίτητος περιορισμός για την εφαρμογή του παραπάνω σχεδιασμού είναι η ταύτιση διαστάσεων των επιπέδων του δικτύου που βρίσκονται στην αρχή και το τέλος της σύνδεσης, ώστε να είναι εφικτή η πράξη της πρόσθεσης. Η συγκεκριμένη σχεδιαστική τεχνική καλείται *μπλοκ με κατάλοιπο* (residual block).

Το *ανεστραμμένο επίπεδο με κατάλοιπο* (inverted residual block) αποτελεί μια τροποποίηση του μπλοκ με κατάλοιπο, λαμβάνοντας υπόψιν την ιδιότητα του επιπέδου γραμμικής συμφόρησης να περιέχει μια αναπαράσταση της χρήσιμης πληροφορίας που διασχίζει το δίκτυο. Έτσι, αντί να τοποθετούμε συνδέσεις καταλοίπου μεταξύ των συνελκτικών πράξεων του δικτύου, επιλέγουμε να τις τοποθετήσουμε μεταξύ των επιπέδων συμφόρησης που παρεμβάλλονται στις συνελκτικές πράξεις. Έτσι, δημιουργούμε μια συνολική αρχιτεκτονική δικτύου που στα επίπεδα συμφόρησης περιέχει τη χρήσιμη πληροφορία που διασχίζει το δίκτυο και μεταξύ δύο τέτοιων σταδίων παρεμβάλλεται η επεξεργασία τους σε υψηλότερες διαστάσεις, για την παραγωγή υψηλότερων χαρακτηριστικών (Εικόνα 3.6).

Ένα μπλοκ συμφόρησης με κατάλοιπο ρυθμίζεται από τις ακόλουθες παραμέτρους (Πίνακα 3.1):

- Επαναλήψεις  $n$ : Το πλήθος επαναλήψεων του σχήματος, συνδεδεμένων σε σειρά. Για κάθε επανάληψη του σχήματος, πλὴν της πρώτης, προστίθεται η δυνατότητα σύνδεσης με κατάλοιπο μεταξύ του επιπέδου επέκτασης και του επιπέδου γραμμικής συμφόρησης.
- Βάθος εισόδου  $c$ : Το βάθος της εξόδου της δομικής μονάδας.
- Παράγοντας επέκτασης  $k$ : Ο παράγοντας επέκτασης της εισόδου κατά το πρώτο στάδιο του σχήματος συμφόρησης, ώστε η διαχωρίσιμη ανά κανάλι συνέλιξη να λάβει είσοδο βάθους  $k \cdot c$ .



Εικόνα 3.6: Αντιπαράθεση του υπολειπόμενου μπλοκ και του ανεστραμμένου υπολειπόμενου μπλοκ (Πηγή [20])

- Βήμα  $s$ : Το βήμα εφαρμογής της διαχωρίσιμης ανά κανάλι συνέλιξης. Σε περίπτωση που το σχήμα επαναλαμβάνεται, το βήμα έχει τιμή  $s$  για το πρώτο σχήμα και τιμή  $s = 1$  για τα επόμενα. Αυτό είναι απαραίτητο για τη διατήρηση ίδιων διαστάσεων στην είσοδο και την έξοδο των μετέπειτα σχημάτων, ώστε να είναι εφικτή η εφαρμογή της σύνδεσης καταλόιπου μεταξύ διαδοχικών επιπέδων.

Είσοδος	Επίπεδο	Έξοδος
$h \times w \times c$	Point-wise Convolution	$h \times w \times (kc)$
$h \times w \times (kc)$	Depth-wise Convolution with stride $s$	$\frac{h}{s} \times \frac{w}{s} \times (kc)$
$\frac{h}{s} \times \frac{w}{s} \times (kc)$	Linear Point-wise Convolution	$\frac{h}{s} \times \frac{w}{s} \times k'$

Πίνακας 3.1: Περιγραφή μπλοκ συμφόρησης με κατάλοιπο (Πηγή [21])

Η βασική υπερπαράμετρος των δικτύων MobileNet είναι ο ‘πολλαπλασιαστής βάθους’ (width multiplier)  $\alpha \in (0, 1]$ , που ρυθμίζει το μέγεθος και την ταχύτητα εκτέλεσης του δικτύου, μεταβάλλοντας πολλαπλασιαστικά το βάθος των επιμέρους συνέλιξεών του. Αν  $\alpha < 1$ , μειώνεται πολλαπλασιαστικά το πλήθος των φίλτρων σε κάθε επίπεδο, αν  $\alpha > 1$ , αυξάνεται πολλαπλασιαστικά το πλήθος των φίλτρων σε κάθε επίπεδο και αν  $\alpha=1$  τότε δεν μεταβάλλεται το πλήθος των φίλτρων σε κάθε επίπεδο.

Η συνολική επίδραση του πολλαπλασιαστή βάθους ‘ $\alpha$ ’ στο υπολογιστικό κόστος της διαχωρίσιμης ανά κανάλι συνέλιξης, υπολογίζεται, τροποποιώντας κατάλληλα την Εξίσωση 3.6, οπότε προκύπτει η Εξίσωση 3.8:

$$H_K \cdot W_K \cdot aM \cdot H_F \cdot W_F + aM \cdot aN \cdot H_F \cdot W_F \quad (3.8)$$

Παρακάτω φαίνεται η συνολική αρχιτεκτονική του μοντέλου MobileNetV2 (Εικόνα 3.7).

Input	Operator	$t$	$c$	$n$	$s$
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	-

Εικόνα 3.7: Αρχιτεκτονική του MobileNetV2 (Πηγή [21])

### 3.3 Βελτιστοποίηση Υπερπαραμέτρων

Τα μοντέλα μηχανικής μάθησης έχουν πολλές ρυθμίσεις που πρέπει να συντονίσουν με στόχο τον έλεγχο της συμπεριφοράς του αλγορίθμου μάθησης. Αυτές οι ρυθμίσεις ονομάζονται *υπερπαραμέτροι*. Οι υπερπαραμέτροι είναι σημεία επιλογής, από τον προγραμματιστή, που επιτρέπουν την προσαρμογή ενός μοντέλου μηχανικής μάθησης για μια συγκεκριμένη εργασία ή σύνολο δεδομένων. Τα μοντέλα μηχανικής μάθησης έχουν επίσης παραμέτρους, οι οποίες είναι οι εσωτερικοί συντελεστές που ορίζονται από την εκπαίδευση ή τη βελτιστοποίηση του μοντέλου σε ένα σύνολο δεδομένων εκπαίδευσης. Οι παράμετροι διαφέρουν από τις υπερπαραμέτρους. Οι παράμετροι μαθαίνονται αυτόματα. Οι υπερπαραμέτροι ρυθμίζονται χειροκίνητα για να βοηθήσουν στην καθοδήγηση της διαδικασίας εκμάθησης.

Συνήθως μια υπερπαραμέτρος έχει μια γνωστή επίδραση σε ένα μοντέλο με τη γενική έννοια, αλλά δεν είναι σαφές πώς να οριστεί καλύτερα μια υπερπαραμέτρος για ένα σύνολο δεδομένων. Επιπλέον, πολλά μοντέλα μηχανικής μάθησης έχουν μια σειρά από υπερπαραμέτρους και μπορεί να αλληλεπιδρούν με μη γραμμικούς τρόπους. Ως εκ τούτου, συχνά απαιτείται η αναζήτηση ενός συνόλου υπερπαραμέτρων που έχουν ως αποτέλεσμα την καλύτερη απόδοση ενός μοντέλου σε ένα σύνολο δεδομένων. Αυτό ονομάζεται *βελτιστοποίηση υπερπαραμέτρων*, αναζήτηση υπερπαραμέτρων ή *hyperparameter tuning*.

Μια διαδικασία βελτιστοποίησης περιλαμβάνει τον καθορισμό ενός χώρου αναζήτησης. Ένα σημείο στο χώρο αναζήτησης είναι ένα διάνυσμα με μια συγκεκριμένη τιμή για κάθε τιμή υπερπαραμέτρου. Ο στόχος της διαδικασίας βελτιστοποίησης είναι να βρεθεί ένα διάνυσμα που έχει ως αποτέλεσμα την καλύτερη απόδοση του μοντέλου μετά την εκμάθηση, δηλαδή

τη μέγιστη ακρίβεια ή/και το ελάχιστο σφάλμα, υπό περιορισμούς (όπως περιορισμό για το χρόνο εκπαίδευσης ή τη μνήμη).

Η διαδικασία βελτιστοποίησης υπερπαραμέτρων αναλύεται στα παρακάτω βήματα :

1. Επιλέγεται ένα σύνολο υπερπαραμέτρων.
2. Δημιουργείται το αντίστοιχο μοντέλο.
3. Εφαρμόζεται το μοντέλο στα δεδομένα εκπαίδευσης και μετράται η τελική απόδοση στα δεδομένα επικύρωσης.
4. Επιλέγεται ο καλύτερος συνδιασμός των υπερπαραμέτρων.
5. Τέλος, μετριέται η απόδοση στα δεδομένα δοκιμής.

Για τη δημιουργία ενός μοντέλου βαθιάς μάθησης, ο προγραμματιστής πρέπει να λάβει πολλές αυθαίρετες αποφάσεις όπως το πλήθος των επιπέδων, το πλήθος των φίλτρων ανά επίπεδο, τις συναρτήσεις ενεργοποίησης κλπ. Αυτές οι παράμετροι σε επίπεδο αρχιτεκτονικής είναι οι υπερπαραμέτροι για να ξεχωρίσουν από τις παραμέτρους ενός μοντέλου, οι οποίες εκπαιδεύονται μέσω backpropagation. Παρακάτω, θα δοθούν οι διάφορες υπερπαραμέτροι που εξερευνήθηκαν.

### 3.3.1 Αλγόριθμοι Βελτιστοποίησης και Ρυθμός Μάθησης

Σε αυτήν την ενότητα, θα συζητηθεί μία σειρά αλγορίθμων οι οποίοι ονομάζονται *Βελτιστοποιητές* (Optimizers). Αυτοί επιδιώκουν να αντιμετωπίσουν την πρόκληση της βελτιστοποίησης των βαθιών μοντέλων κατά την εκπαίδευση των συνελκτικών νευρωνικών δικτύων. Αυτό το επιτυγχάνουν προσαρμόζοντας το *ρυθμό μάθησης* (learning rate) για κάθε παράμετρο του μοντέλου, έτσι ώστε να ελαχιστοποιηθεί όσο το δυνατόν περισσότερο η συνάρτηση κόστους. Ο ρυθμός μάθησης είναι μία υπερπαραμέτρος βελτιστοποίησης που ελέγχει το μέγεθος του βήματος που λαμβάνουν οι παράμετροι προς την κατεύθυνση της κλίσης. Παρακάτω παρουσιάζονται οι βελτιστοποιητές που χρησιμοποιήθηκαν στα πειράματα της παρούσας διπλωματικής εργασίας.

- Adam: Ο “Adam” παρουσιάστηκε πρώτη φορά από τον Diederik Kingma και τον Jimmy Ba το 2014 [97]. Είναι μια επέκταση της στοχαστικής κατάβασης κλίσης και το όνομα του προέρχεται από την ‘προσαρμοστική εκτίμηση ροπής’ (Adaptive moment estimation). Όμως, η στοχαστική κατάβαση κλίσης διατηρεί τον ίδιο ρυθμό μάθησης  $\alpha$  για όλες τις ανανεώσεις βαρών και δεν αλλάζει κατά τη διάρκεια της εκπαίδευσης ενώ στον Adam κάθε βάρος ανανεώνεται με διαφορετικό ρυθμό μάθησης και μάλιστα αυτό μπορεί να αλλάξει κατά τη διάρκεια της εκπαίδευσης, δηλαδή για κάθε νέο δείγμα εκπαίδευσης. Επίσης, ο Adam συνδιάζει τα πλεονεκτήματα δύο άλλων βελτιστοποιητών, του “AdaGrad” και του “RMSProp”. Ο αλγόριθμος προσαρμοστικής κλίσης (AdaGrad) διατηρεί ένα ρυθμό μάθησης ανά παράμετρο και βελτιώνει την απόδοση σε προβλήματα με αραιές διαβαθμίσεις (π.χ. προβλήματα φυσικής γλώσσας). Ο Root Mean Square Propagation (RMSProp) διατηρεί επίσης ρυθμούς μάθησης ανά παράμετρο που προσαρμόζονται με βάση τον μέσο όρο των πρόσφατων μεγεθών των κλίσεων για το βάρος.



Αυτό σημαίνει ότι ο αλγόριθμος τα πάει καλά σε μη σταθερά δεδομένα (π.χ. θορυβώδη). Ο Adam αντιθέτως, κάνει χρήση του μέσου όρου και των δεύτερων ροπών των κλίσεων (μη κεντρική διακύμανση). Πιο συγκεκριμένα, εφαρμόζει τον εκθετικό κινητό μέσο όρο της κλίσης ( $m_t$ ) και της τετραγωνικής κλίσης ( $v_t$ ) και οι υπερπαραμέτροι  $\beta_1, \beta_2 \in [0, 1)$  ελέγχουν τους ρυθμούς εκθετικής αποσύνθεσης αυτών των κινητών μέσων. Οι κινούμενοι μέσοι όροι αρχικοποιούνται ως 0, οδηγώντας σε εκτιμήσεις κλίσης που είναι μεροληπτικές γύρω στο 0, ειδικά κατά τα αρχικά βήματα. Οι Εξισώσεις 3.9-3.13 εκφράζουν όλα τα παραπάνω μεγέθη.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (3.9)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (3.10)$$

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3.11)$$

$$\widehat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3.12)$$

$$\theta \leftarrow \theta - \frac{a \cdot \widehat{m}_t}{\sqrt{\widehat{v}_t} + \varepsilon} \quad (3.13)$$

,όπου  $g_t$  το gradient.

- **Nadam:** Ο “Nadam” παρουσιάστηκε πρώτη φορά το 2015 από τον Dozat [98] και ονομάστηκε έτσι από τη φράση “Nesterov-accelerated Adaptive Moment Estimation”. Είναι μια επέκταση του Adam συνδυάζοντας τον με τον βελτιστοποιητή “NAG”, ο οποίος βελτιώνει την κατάβαση πιο γρήγορα. Λειτουργεί ελαφρώς καλύτερα από την τυπική βελτιστοποίηση με ορμή. Ο Nadam χρησιμοποιείται για θορυβώδεις κλίσεις ή για κλίσεις με υψηλές καμπυλότητες. Η διαδικασία εκμάθησης επιταχύνεται αθροίζοντας την εκθετική μείωση των κινητών μέσων για την προηγούμενη και την τρέχουσα κλίση. Τα βάρη ενημερώνονται με βάση την Εξίσωση 3.14 και στις Εξισώσεις (3.15-3.17) φαίνονται πως υπολογίζονται οι άλλες μεταβλητές.

$$w_t^i = w_{t-1}^i - \frac{a \cdot \overline{m}_t}{\sqrt{u_t} + \varepsilon} \quad (3.14)$$

όπου:

$$\overline{m}_t = \beta_1^{t+1} \cdot \widehat{m}_t + (1 - \beta_1^t) \cdot \widehat{g}_t \quad (3.15)$$

$$\widehat{m}_t = \frac{m_t}{1 - \prod_{i=1}^t \beta_1^i} \quad (3.16)$$

$$\widehat{g}_t = \frac{g_t}{1 - \prod_{i=1}^t \beta_1^i} \quad (3.17)$$

### 3.3.2 Μέγεθος δέσμης

Το μέγεθος δέσμης (batch size) είναι ένας ακέραιος αριθμός δειγμάτων που υποβάλλονται σε επεξεργασία πριν από την ενημέρωση του μοντέλου, δηλαδή μετά από σύνολα δειγμάτων εκπαίδευσης θα ανανεώνονται οι κλίσεις. Το μέγεθος δέσμης πρέπει να είναι μεγαλύτερο ή ίσο με 1 και μικρότερο ή ίσο με τον αριθμό των δειγμάτων, στο σύνολο δεδομένων εκπαίδευσης.

### 3.3.3 Εποχές

Ο αριθμός των εποχών (epochs) είναι το πλήθος των πλήρων περασμάτων από το σύνολο δεδομένων εκπαίδευσης. Ο αριθμός των εποχών ορίζεται σε μια ακέραια τιμή μεταξύ 1 και άπειρου. Ο αλγόριθμος μπορεί να εκτελεστεί για ένα συγκεκριμένο διάστημα, ορίζοντας ένα σταθερό αριθμό εποχών, αλλά και να σταματήσει χρησιμοποιώντας άλλα κριτήρια, όπως μια αλλαγή στο σφάλμα του μοντέλου με την πάροδο του χρόνου.

### 3.3.4 Πλήθος των Πλήρως διασυνδεδεμένων επιπέδων

Το τελευταίο, πλήρως διασυνδεδεμένο, επίπεδο που χρησιμοποιείται στα δίκτυα της εργασίας, έχει πάντα δύο κόμβους καθώς το πρόβλημα δυαδικής ταξινόμησης. Δοκιμάστηκαν οι παρακάτω αρχιτεκτονικές:

- Ένα κρυφό επίπεδο με 1024 νευρώνες.
- Δύο κρυφά επίπεδα με το πρώτο να έχει 1024 νευρώνες και το δεύτερο 512 νευρώνες.

### 3.3.5 Αρχικοποιητές πυρήνα

Οι *αρχικοποιητές πυρήνα* (kernel initializers) χρησιμοποιούνται για την στατιστική αρχικοποίηση των παραμέτρων του μοντέλου. Στο πλαίσιο της παρούσας εργασίας έγινε αρχικοποίηση τόσο με βάση την κανονική κατανομή όσο και με την περικομμένα κανονική κατανομή (με μέση τιμή 0 και τυπική απόκλιση που υπολογίζεται σύμφωνα με την Εξίσωση 3.18).

$$standard\_deviation = \sqrt{\frac{2}{fan\_in + fan\_out}} \quad (3.18)$$

,όπου fan\_in είναι ο αριθμός των μονάδων εισόδου στον τανυστή βάρους και fan\_out είναι ο αριθμός των μονάδων εξόδου στον τανυστή βάρους.

### 3.3.6 Ποσοστό εγκατάλειψης

Η εγκατάλειψη εισάγεται με το Dropout Layer. Στα δίκτυα που υλοποιούνται στην προκείμενη εργασία εισάγεται κυρίως πριν το Flatten Layer και πριν/μετά τα πλήρως διασυνδεδεμένα επίπεδα γιατί εκεί βρίσκονται οι περισσότεροι κόμβοι-παράμετροι.

## Κεφάλαιο 4

# Σύνολο Δεδομένων και Εργαλεία

---

Στο κεφάλαιο αυτό παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων μηχανικής μάθησης καθώς και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν για την υλοποίησή τους.

### 4.1 Δεδομένα

Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε το σύνολο δεδομένων εναέριων εικόνων FLAME (Fire Luminosity Airborne-based Machine learning Evaluation) [22]. Αυτό ανακτήθηκε χρησιμοποιώντας drones κατά τη διάρκεια μιας προδιαγεγραμμένης καύσης πασσάλων, στη Βόρεια Αριζόνα των ΗΠΑ. Αποτελείται από διαφορετικά αποθετήρια, συμπεριλαμβανομένων ακατέργαστων εναέριων βίντεο που καταγράφηκαν από κάμερες drones και επίσης ακατέργαστων πλανών από θερμικό χάρτη που καταγράφηκαν από μια υπέρυθρη θερμική κάμερα.

Περιλαμβάνει εικόνες, οι οποίες είναι περιορισμένης ανάλυσης, αφού έχουν τραβηχτεί από μεγάλο ύψος. Επίσης, το σύνολο εκπαίδευσης, όπως θα δούμε και στις εικόνες παρακάτω, περιέχει και εικόνες επισημασμένες με "Φωτιά" στις οποίες δεν φαίνεται ξεκάθαρα η φωτιά ούτε με το ανθρώπινο μάτι. Περιέχουν κάπνο όμως, που σε αντίστοιχες εικόνες του συνόλου δοκιμών υπάρχουν εικόνες επισημασμένες "Χωρίς Φωτιά" που περιέχουν ομίχλη ενώ θα μπορούσε να είναι καπνός. Το σύστημα μπορεί εύκολα να μπερδευτεί όταν δεν υπάρχει έντονα το χρώμα της φωτιάς ή όταν ο καπνός δεν είναι αρκετά σκουρόχρωμος. Αυτό δικαιολογεί και την πιθανή απόκλιση-σφάλμα στα δίκτυα που εκπαιδεύονται.

Για τη προσέγγιση νευρωνικών δικτύων και ταξινόμηση πυρκαγιάς, χρησιμοποιήθηκε το έβδομο και το όγδοο αποθετήριο. Το έβδομο αποτελείται από 39.375 πλαίσια μεγέθους  $254 \times 254$  εικονοστοιχείων, επισημασμένα ως "Fire" vs "Non-Fire" για τη φάση εκπαίδευσης. Το μέγεθος αυτού του αποθετηρίου είναι 1,3 GB και η μορφή είναι JPEG. Επίσης, το όγδοο αποτελείται από άλλα 8.617 πλαίσια μεγέθους  $254 \times 254$  εικονοστοιχεία, επισημασμένα ως "Fire" vs "Non-Fire" για τη φάση δοκιμών. Το μέγεθος αυτού του αποθετηρίου είναι 301 MB και η μορφή είναι JPEG (Εικόνες 4.1-4.2).

Χωρίσαμε το σύνολο δεδομένων εκπαίδευσης σε δεδομένα εκπαίδευσης και δεδομένα επικύρωσης, σε ποσοστό 80% και 20%, αντίστοιχα.



Εικόνα 4.1: Παραδείγματα εικόνων φωτιάς από το training set (Πηγή [22])

## 4.2 Εργαλεία

Οι αρχιτεκτονικές που μελετήθηκαν προηγουμένως έχουν υλοποιηθεί στη γλώσσα προγραμματισμού Python 3. Η Python είναι μια γλώσσα υψηλού επιπέδου, κατάλληλη για εφαρμογές μηχανικής μάθησης και κατ' επέκταση για νευρωνικά δίκτυα. Είναι μια γλώσσα που χαρακτηρίζεται από απλότητα, συνέπεια, ανεξαρτησία πλατφόρμας, ευρεία κοινότητα και εξαιρετικές βιβλιοθήκες.

Παρακάτω περιγράφονται μερικές βιβλιοθήκες που χρησιμοποιήθηκαν στην υλοποίηση.

- **Tensorflow:** Η Tensorflow είναι μια βιβλιοθήκη ανοικτού κώδικα, που χρησιμοποιείται κυρίως για υλοποιήσεις νευρωνικών δικτύων. Αναπτύχθηκε από την Google Brain, την ομάδα τεχνητής νοημοσύνης της Google [99]. Η Tensorflow λειτουργεί σε large scale και σε ετερογενή περιβάλλοντα. Σκοπός της δημιουργίας της ήταν η διευκόλυνση διαδικασιών όπως ο προγραμματισμός ροής δεδομένων, σύμφωνα με το οποίο ένα



Εικόνα 4.2: Παραδείγματα εικόνων χωρίς φωτιά από το test set (Πηγή [22])

πρόγραμμα μοντελοποιείται σαν ένας προσανατολισμένος γράφος πράξεων πάνω σε δομές δεδομένων. Το όνομα της προκύπτει από τους πολυδιάστατους πίνακες δεδομένων, που λέγονται τανυστές (tensors), στους οποίους γίνονται οι διαδικασίες που εκτελούν τα νευρωνικά δίκτυα. Η TF μπορεί να εκμεταλλευτεί την υπολογιστική ισχύ πολλαπλών επεξεργαστών και καρτών γραφικών GPU, στις οποίες παρέχει δυνατότητα ανάπτυξης.

- cuDNN: Η cuDNN (NVIDIA CUDA Deep Neural Network) είναι μια βιβλιοθήκη με επιτάχυνση GPU για εφαρμογές βαθιών νευρωνικών δικτύων. Παρέχει εξαιρετικά συντονισμένες υλοποιήσεις ρουτινών που προκύπτουν συχνά σε εφαρμογές DNN [100].
- Keras: Η Keras είναι μια βιβλιοθήκη ανοικτού κώδικα, κατάλληλη για νευρωνικά δίκτυα [101]. Λειτουργεί ως διεπαφή για τη βιβλιοθήκη Tensorflow ή άλλες βιβλιοθήκες για νευρωνικά δίκτυα, το Theano [102] και το CNTK (The Microsoft Cognitive Toolkit) [103]. Η Keras είναι πολύ φιλική στο χρήστη και γρήγορα μπορεί κάποιος

να αναπτύξει και υλοποιήσει μια ιδέα. Είναι δυνατόν να εξαχθούν μοντέλα Keras σε JavaScript για να εκτελεστούν απευθείας σε πρόγραμμα περιήγησης ή εναλλακτικά σε TF Lite για εκτέλεση σε iOS, Android και ενσωματωμένες συσκευές. Χρησιμοποιείται από οργανισμούς και εταιρίες όπως η NASA, το CERN και το Youtube.

- **Scikit-Learn:** Η Scikit-learn (ή sklearn) είναι επίσης μια βιβλιοθήκη μηχανικής μάθησης ανοικτού κώδικα της γλώσσας Python [104]. Διαθέτει αλγόριθμους για ταξινόμηση (classification), παλινδρόμηση (regression), μηχανές διανυσμάτων υποστήριξης (SVM), αλλά και αλγόριθμους μη επιβλεπόμενης μάθησης όπως k-means και DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Έχει σχεδιαστεί για να συνεργάζεται με τις αριθμητικές και επιστημονικές βιβλιοθήκες της Python, NumPy και SciPy.
- **NumPy:** Η NumPy, μια ακόμα βιβλιοθήκη ανοικτού κώδικα, είναι το θεμελιώδες πακέτο για επιστημονικούς υπολογισμούς με Python. Πρωτοεμφανίστηκε ως “numeric” το 1995 και το 2005 έλαβε τη σημερινή της ονομασία. Κάνει εφικτούς υπολογισμούς με πολυδιάστατους πίνακες χρησιμοποιώντας στο υπόβαθρο ρουτίνες γραμμένες σε C [105].
- **SciPy:** Η ανοικτού κώδικα βιβλιοθήκη SciPy περιέχει μονάδες για βελτιστοποίηση, γραμμική άλγεβρα και άλλα εργαλεία επιστημονικών υπολογισμών. Η κύρια δομή δεδομένων της είναι ένας πολυδιάστατος πίνακας που της παρέχεται από τη βιβλιοθήκη NumPy. Επίσης η NumPy παρέχει συναρτήσεις γραμμικής άλγεβρας, μετασχηματισμούς Fourier και άλλες μαθηματικές δομές [106].
- **Matplotlib:** Η Matplotlib είναι μια βιβλιοθήκη σχεδίασης και οπτικοποίησης για τη γλώσσα προγραμματισμού Python. Το Pyplot είναι μια λειτουργική μονάδα Matplotlib που παρέχει μια διεπαφή παρόμοια με τη MATLAB. Η Matplotlib έχει σχεδιαστεί για να είναι τόσο χρησιμοποιήσιμη όσο τη MATLAB, γι’αυτό και πήρε το όνομα της από αυτήν, με δυνατότητα χρήσης Python και το πλεονέκτημα ότι είναι δωρεάν και ανοικτού κώδικα [107].

Επιπλέον, η εκπόνηση του υπολογιστικού μέρους της εργασίας έγινε στο περιβάλλον του λειτουργικού συστήματος Windows 10. Η επιτάχυνση της εκτέλεσης αριθμητικών πράξεων, και γενικότερα η εκπαίδευση των μοντέλων, έγινε εφικτή με τη χρήση κάρτας γραφικών GPU NVIDIA GeForce GTX 1050 Ti με 4GB memory size.

## Κεφάλαιο 5

# Πρακτική υλοποίηση και αποτελέσματα

Στο κεφάλαιο αυτό περιγράφεται η πειραματική διαδικασία, με βάση τη μελέτη που έγινε για την υλοποίηση των μοντέλων. Αρχικά, περιγράφονται οι τρόποι αξιολόγησης της απόδοσης των μοντέλων. Ύστερα, περιγράφονται αναλυτικά οι υλοποιήσεις των δυο μοντέλων που έγιναν με μεταφορά μάθησης καθώς και οι υλοποιήσεις των δυο μοντέλων που έγιναν από την αρχή.

### 5.1 Αξιολόγηση Απόδοσης

Για να ποσοτικοποιηθεί η επιτυχία της εκπαίδευσης ενός νευρωνικού μοντέλου, χρησιμοποιούνται μετρικές αξιολόγησης στο σύνολο ελέγχου (test set). Σ' αυτήν την εργασία θα μελετηθούν οι παρακάτω βασικές μετρικές για όλες τις βέλτιστες υλοποιήσεις των 4 μοντέλων που προέκυψαν.

- Πίνακας Σύγχυσης (Confusion Matrix): Πρόκειται για έναν τετραγωνικό πίνακα μεγέθους  $N \times N$ , που δίνει συγκεντρωτικά τα αποτελέσματα πρόβλεψης σε ένα πρόβλημα ταξινόμησης (Εικόνα 5.1). Ο αριθμός των σωστών και εσφαλμένων προβλέψεων συνοψίζεται με τιμές μέτρησης και αναλύεται ανά τάξη. Στόχος είναι ο πίνακας σύγχυσης να έχει μηδενικά στα στοιχεία εκτός της κύριας διαγωνίου.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Εικόνα 5.1: Πίνακας σύγχυσης δύο κλάσεων (Πηγή [23])

- Πισότητα (Accuracy): Είναι μια από τις βασικότερες μετρικές αξιολογήσεις ενός μοντέλου. Ορίζεται, όπως φαίνεται και στην Εξίσωση 5.1, ως το κλάσμα των ορθών προβλέψεων προς το πλήθος των συνολικών εκτιμήσεων που πραγματοποιήθηκαν.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.1)$$

Όμως, όταν το σύνολο δεδομένων δεν είναι ισορροπημένο, δηλαδή το πλήθος των δειγμάτων είναι άνισο στις κατηγορίες, πρέπει να λαμβάνονται υπόψη κι άλλες μετρικές όπως είναι το Precision, Recall, Sensitivity, F1-score.

- Ανάκληση (Recall): Στην προκείμενη εργασία λαμβάνεται υπόψη εκτός από την πισότητα και η μετρική της ανάκλησης. Αυτό γίνεται διότι έχουμε μεγαλύτερο κόστος όταν έχουμε περισσότερα ψευδώς αρνητικά. Είναι εμφανές ότι αν ένα σύστημα κατηγοριοποιήσει μια εικόνα που έχει φωτιά στην λάθος κατηγορία No-Fire, οι επιχειρήσεις κατάσβεσης θα καθυστερήσουν να μεταβούν στο σημείο πριν εξαπλωθεί. Ορίζεται και ως True Positive Rate και η μορφή της δίνεται στην Εξίσωση 5.2.

$$Recall = \frac{TP}{TP + FN} \quad (5.2)$$

- F1-score: Αξίζει να γίνει μια αναφορά και στη μετρική του F1-score. Ορίζεται ως ο σταθμισμένος μέσος όρος της Ακρίβειας (Precision) και της Ανάκλησης (Recall). Είναι μια πολύ πιο χρήσιμη μετρική για άνισες σε πλήθος κατηγορίες από αυτή της ακρίβειας αφού δίνεται βάση και στα ψευδώς θετικά αλλά και στα ψευδώς αρνητικά αποτελέσματα. Μαθηματικά ορίζεται όπως φαίνεται παρακάτω στις Εξισώσεις 5.3-5.4:

$$F1 - score = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall} \quad (5.3)$$

Η διαφορετικά ως:

$$F1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (5.4)$$

Σημειώνουμε ότι αν έχουμε τον πίνακα σύγκρισης του μοντέλου, είναι πολύ εύκολο με τις παραπάνω εξισώσεις να εξαγάγουμε τις υπόλοιπες μετρικές.

## 5.2 Υλοποιήσεις με μεταφορά μαθησης

Οι μέθοδοι ταξινόμησης που εφαρμόστηκαν είναι αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων. Πιο συγκεκριμένα, υλοποιήθηκαν διάφορα μοντέλα Keras από το αποθετήριο που έχει κατασκευαστεί και είναι προσβάσιμο από το Tensorflow. Τα "Keras models" είναι μοντέλα βαθιάς μάθησης που διατίθενται μαζί με προεκπαιδευμένα βάρη. Αυτά τα μοντέλα μπορούν να χρησιμοποιηθούν για πρόβλεψη, εξαγωγή χαρακτηριστικών και μεταφορά μάθησης. Τα βάρη μεταφορτώνονται αυτόματα κατά την προετοιμασία ενός μοντέλου. Κατά



την εγκατάσταση, τα μοντέλα θα κατασκευαστούν σύμφωνα με τη μορφή δεδομένων εικόνας που έχει οριστεί στο αρχείο διαμόρφωσης Keras.

Προτιμήθηκαν να χρησιμοποιηθούν μοντέλα πιο σύγχρονα, που είχαν χρησιμοποιηθεί και σε άλλες συναφείς εργασίες (που περιγράφηκαν στο Κεφάλαιο 4) και μοντέλα που αποτελούνται από λίγες παραμέτρους. Δοκιμάστηκαν τα EfficientNetB0 με 5,3 εκατομμύρια αρχικές παραμέτρους, MobileNetV2 με 3,5 εκατομμύρια αρχικές παραμέτρους, Xception με 22,9 εκατομμύρια αρχικές παραμέτρους, InceptionV3 με 23,9 εκατομμύρια αρχικές παραμέτρους, ResNet50 με 25,6 εκατομμύρια αρχικές παραμέτρους. Όμως, το EfficientNetB0 είχε ακρίβεια κοντά στο 0,6 και χαμηλότερα ενώ τα InceptionV3, ResNet50 είχαν ακρίβεια κάτω του 0,7. Επομένως, επιλέχθηκαν να μελετηθούν τα Xception και MobileNetV2.

Για την εκπαίδευσή τους χρησιμοποιήθηκαν τα δεδομένα εκπαίδευσης (training data) και τα δεδομένα ελέγχου (test data) όπως έχουν χωριστεί από το σύνολο δεδομένων μας. Επίσης, όπως έχει ήδη αναφερθεί, τα δεδομένα εκπαίδευσης χωρίστηκαν και το 20% (τυχαίες εικόνες) ορίστηκαν ως δεδομένα επικύρωσης.

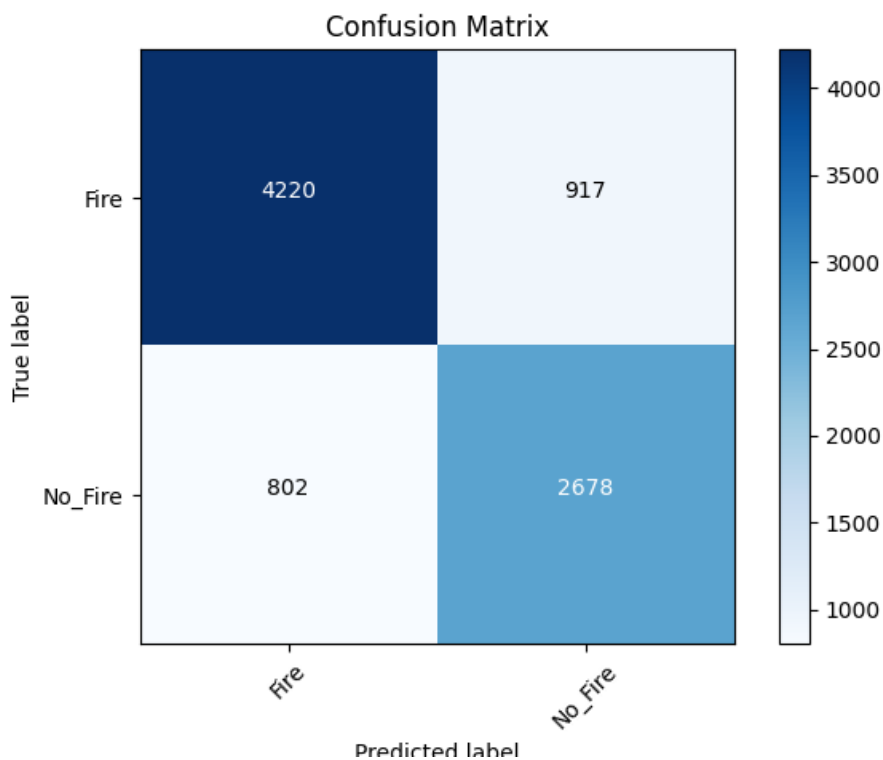
Ένα νευρωνικό δίκτυο CNN, έχει αρκετές παραμέτρους που πρέπει να προσδιοριστούν πριν γίνει η εκπαίδευσή του. Κάποιες από αυτές είναι, το πόσα πυκνά επίπεδα θα χρησιμοποιηθούν και πόσες μονάδες θα έχει το κάθε επίπεδο, ο αλγόριθμος βελτιστοποίησης και το μέγεθος δέσμης. Για τον καθορισμό των υπερπαραμέτρων του μοντέλου έγιναν δοκιμές για την επιλογή των βέλτιστων. Στον παρακάτω Πίνακα 5.1 φαίνονται οι δοκιμές με τους συνδυασμούς των υπερπαραμέτρων που πραγματοποιήθηκαν για την αρχιτεκτονική Xception. Μια υπερπαραμέτρος σε συνδυασμό με τις άλλες προκαλεί είτε μια μικρή μείωση ή μια μικρή βελτίωση στην ακρίβεια του μοντέλου.

Υπερπαραμέτροι	Συνδυασμοί	Βέλτιστα
Hidden FC Layers	[1, 2]	2
FC neurons	[1024, 512]	1024 + 512
Optimizer	[Adam, Nadam]	Nadam
Dropout	[none, 0.2, 0.4, 0.5]	0.5
Batch size	[10, 32]	32
Kernel initializer	[glorot uniform, normal, glorot normal]	normal
Learning rate	[1e-3, 1e-4]	1e-4
Epochs	[10, early stopping]	25

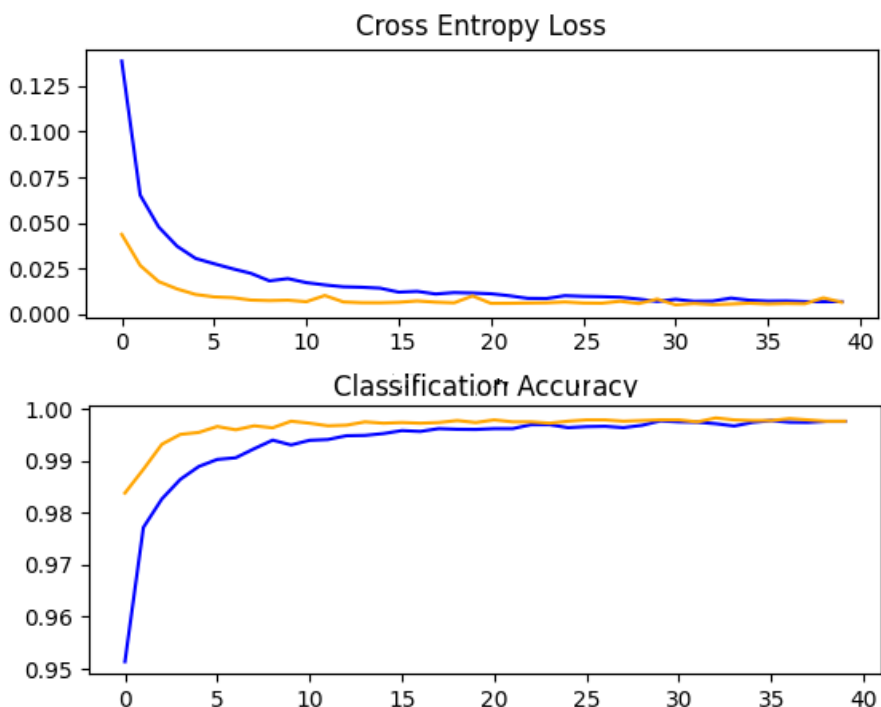
**Πίνακας 5.1:** Πίνακας υπερπαραμέτρων για τη μεταφορά μάθησης με πρώτο μοντέλο Keras – **Xception**

Ακολουθεί ο πίνακας σύγχυσης (Εικόνα 5.2) και το διάγραμμα της απώλειας και της πιστότητας ως προς τις εποχές (Εικόνα 5.3). Στο διάγραμμα έχουμε για τα δεδομένα εκπαίδευσης το μπλε χρώμα και για τα δεδομένα επικύρωσης το πορτοκαλί χρώμα.

Χρησιμοποιήθηκε ο πρόορος τερματισμός με 'υπομονή' 5 ή/και περισσότερων εποχών ώστε να σταματά η εκπαίδευση όταν η μετρική του σφάλματος, στο δείγμα επικύρωσης,



Εικόνα 5.2: Πίνακας σύγχυσης πρώτου μοντέλου



Εικόνα 5.3: Διάγραμμα του πρώτου μοντέλου για την πισιότητα και την απώλεια στα σύνολα εκπαίδευσης και επικύρωσης ως προς τις εποχές

σταματήσει να μειώνεται. Τα callbacks χρησιμοποιούνται για να εκτελεστούν ενέργειες σε διάφορα στάδια της εκπαίδευσης (π.χ. στην αρχή ή στο τέλος μιας εποχής, πριν ή μετά

από μια μεμονωμένη δέσμη, κ.λπ.). Άλλοι λόγοι για να χρησιμοποιηθούν τα callbacks είναι για την εγγραφή καταγραφής του TensorBoard μετά από κάθε παρτίδα εκπαίδευσης, για την παρακολούθηση των μετρήσεων, την αποθήκευση του μοντέλου στο δίσκο με μια περιοδικότητα και την παρατήρηση των εσωτερικών καταστάσεων και τα στατιστικά ενός μοντέλου κατά τη διάρκεια της εκπαίδευσης.

Στην εργασία μας χρησιμοποιήθηκε και το Model Checkpoint ώστε να αποθηκεύεται κατά τη διάρκεια της εκπαίδευσης, το καλύτερο σε επίδοση μοντέλο μέχρι εκείνη την εποχή. Με αυτόν τον τρόπο καταφέρνουμε μεταγενέστερα να φορτώσουμε το αποθηκευμένο μοντέλο και να δούμε την απόδοση στο σύνολο δοκιμών.

Στον παρακάτω Πίνακα 5.2 φαίνονται οι δοκιμές με τους συνδιασμούς των υπερπαραμέτρων που πραγματοποιήθηκαν για την αρχιτεκτονική MobileNetV2. Παρατηρούμε ότι εκτός από τους παραπάνω συνδιασμούς, που έγιναν για το Xception, δοκιμάστηκε ακόμη και η υπερπαραμέτρος alpha του μοντέλου. Η default τιμή της είναι 1 και δοκιμάστηκε ακόμη και η τιμή 0.75 και η τιμή 1.3.

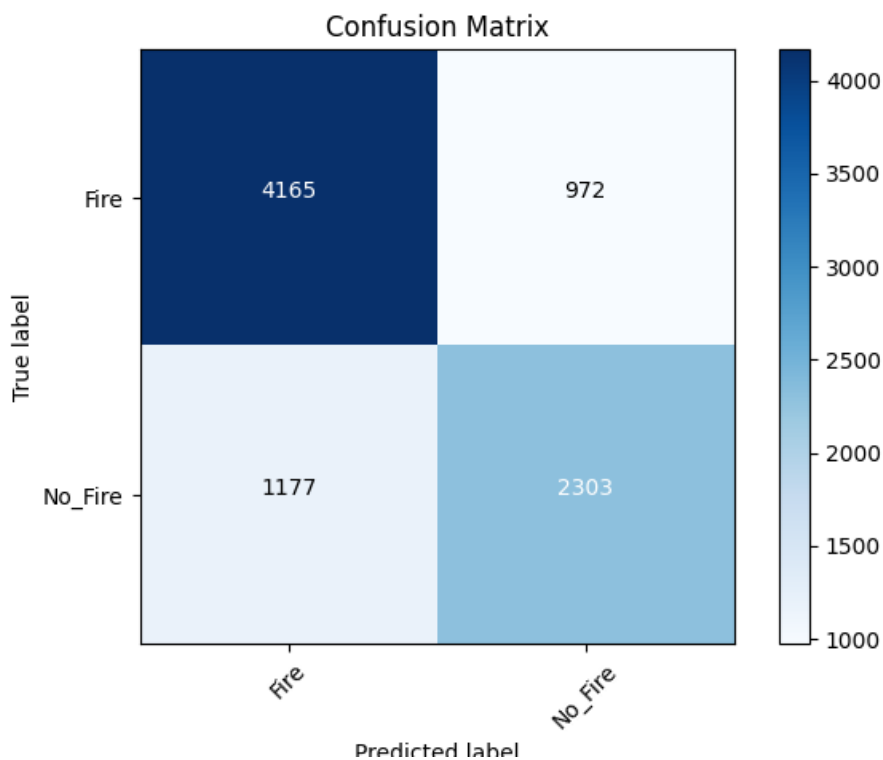
Υπερπαραμέτροι	Συνδιασμοί	Βέλτιστα
Hidden FC Layers	[1, 2]	1
FC neurons	[1024, 512]	1024
Optimizer	[Adam, Nadam]	Nadam
Dropout	[none, 0.2, 0.4, 0.5]	0.5
alpha	[1, 0.75, 1.3]	1
Batch size	[10, 32]	32
Kernel initializer	[glorot uniform, normal, glorot normal]	glorot normal
Learning rate	[1e-3, 1e-4]	1e-3
Epochs	[10, early stopping]	13

**Πίνακας 5.2:** Πίνακας υπερπαραμέτρων για τη μάθηση μέσω μεταφοράς με το δεύτερο μοντέλο Keras – **MobileNetV2**

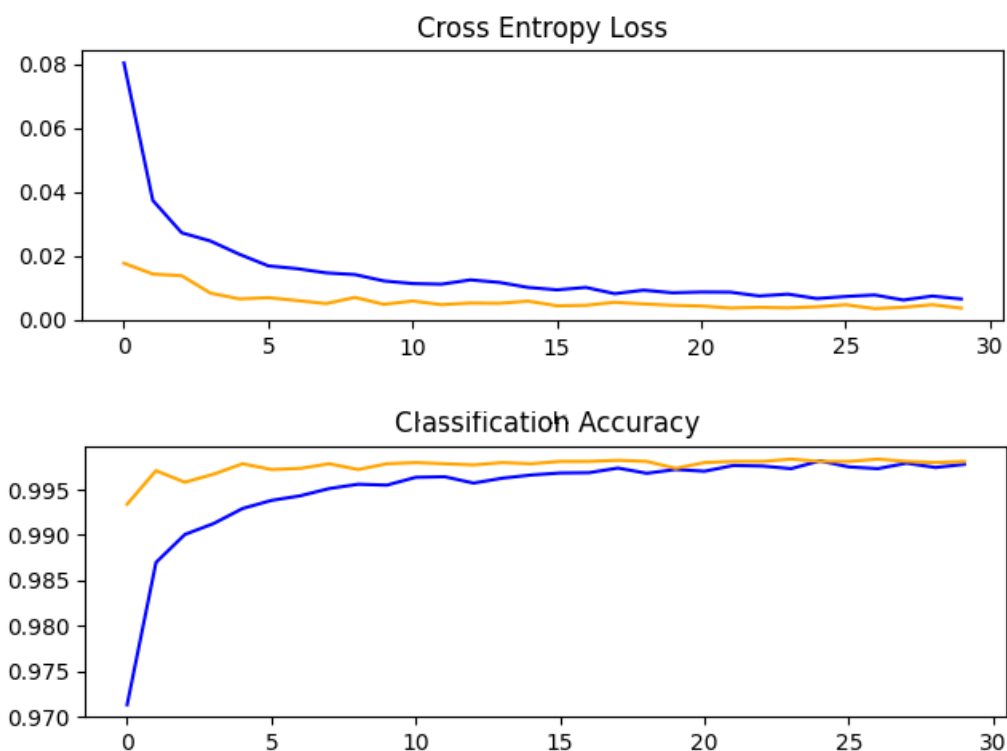
Παρακάτω, παρατίθεται ο πίνακας σύγχυσης (Εικόνα 5.4) καθώς και το διάγραμμα της απώλειας και της πιστότητας για τα δεδομένα εκπαίδευσης (μπλε χρώμα) και τα δεδομένα επικύρωσης (πορτοκαλί χρώμα) σε συνάρτηση με τις εποχές (Εικόνα 5.5).

Ένας άλλος συνδιασμός που θα μπορούσε να επιλεγεί ως βέλτιστος ήταν η υλοποίηση με alpha=1.3, 1 hidden layer με 1024 νευρώνες, Adam optimizer, Dropout=0.5, Batch size=32, kernel initializer = normal, learning rate=1e-4, epochs=30. Η πιστότητα είναι ελάχιστα καλύτερη στο 0,754, η ανάκληση όμως μειώθηκε στο 0,76 όπως και το F1-score στο 0,787. Οι παράμετροι είναι πολύ περισσότεροι και συγκεκριμένα 5.473.058. Άρα, δεν επιλέγεται γιατί υπάρχει αύξηση μόνο στην ακρίβεια και όχι στις άλλες 2 μετρικές που μας ενδιαφέρουν και επιπλέον έχουμε περίπου 2 εκατομμύρια περισσότερες παραμέτρους στην περίπτωση αυτή.

Οι μετρικές αξιολόγησης για το μοντέλο που επιλέχθηκε παρουσιάζονται στο τέλος του



Εικόνα 5.4: Πίνακας σύγχυσης δεύτερου μοντέλου

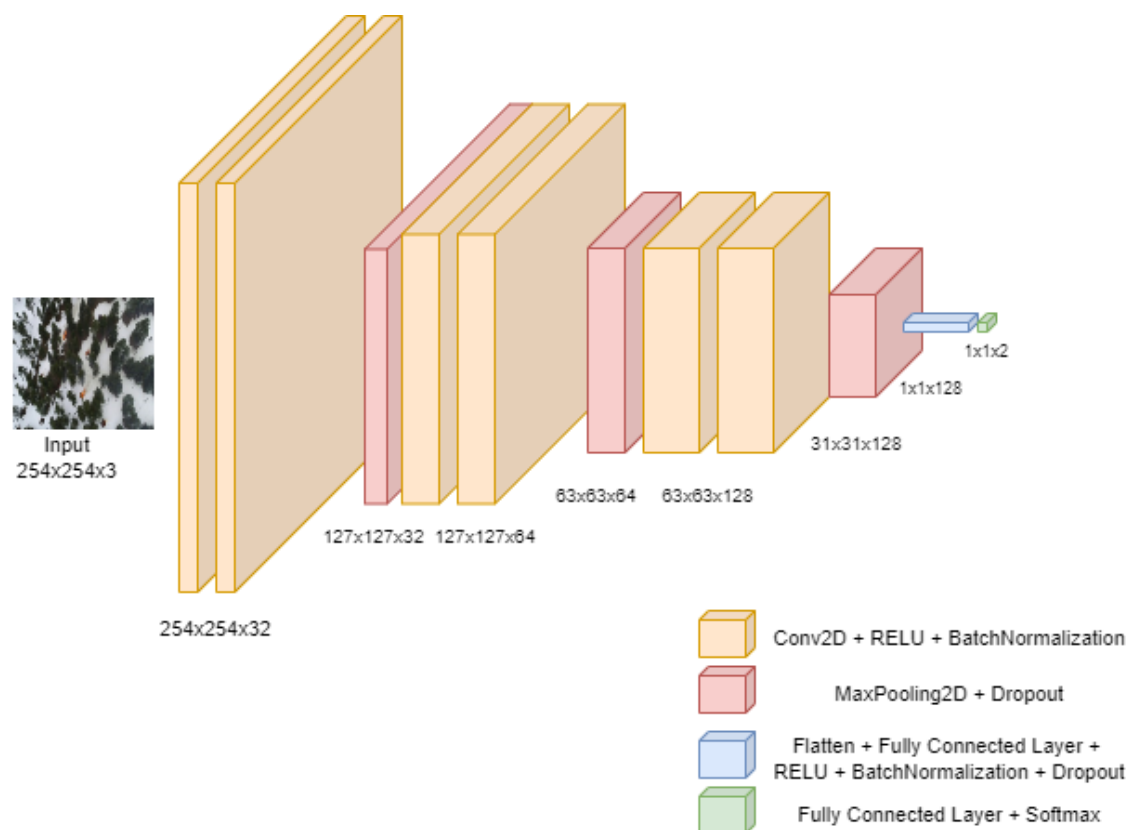


Εικόνα 5.5: Διάγραμμα του τρίτου μοντέλου για την πιστότητα και την απώλεια στα σύνολα εκπαίδευσης και επικύρωσης ως προς τις εποχές

κεφαλαίου, στον συγκεντρωτικό Πίνακα 5.3.

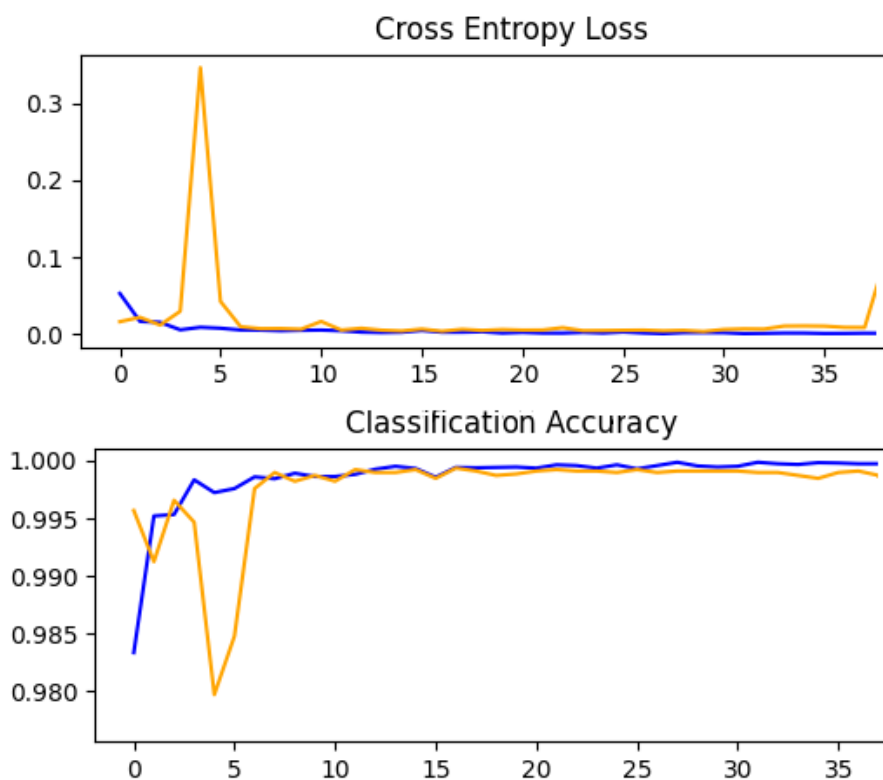
### 5.3 Υλοποιήσεις με κατασκευή από την αρχή

Εκτός από τις υλοποιήσεις μεταφοράς μάθησης που αναλύθηκαν παραπάνω, μελετήθηκαν και δύο μοντέλα, που κατασκευάστηκαν και εκπαιδεύτηκαν από την αρχή, στο πλαίσιο της παρούσας διπλωματικής εργασίας.

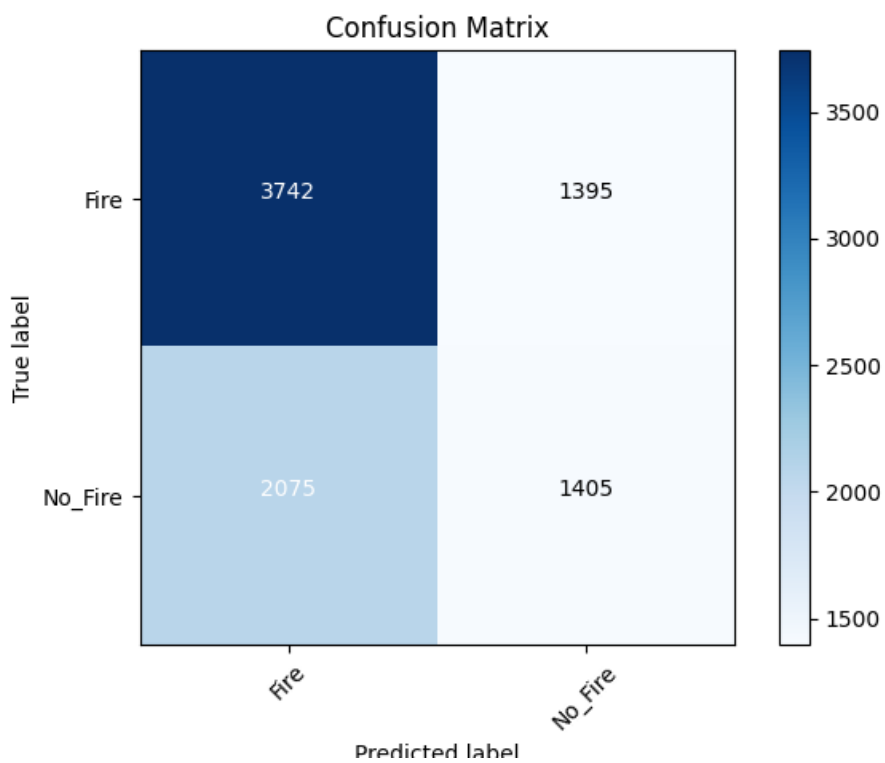


Εικόνα 5.6: Τρίτο μοντέλο

Το τρίτο μοντέλο, που η αρχιτεκτονική του απεικονίζεται στην Εικόνα 5.6, περιλαμβάνει τη στοίβαξη συνεκτικών στρώματων με μικρά φίλτρα  $3 \times 3$  που ανά 2 ακολουθούνται από ένα στρώμα υποδειγματοληψίας μεγίστου. Μαζί, αυτά τα επίπεδα σχηματίζουν ένα μπλοκ και αυτά τα μπλοκ επαναλαμβάνονται με τον αριθμό των φίλτρων σε κάθε μπλοκ να αυξάνεται με το βάθος του δικτύου. Πιο συγκεκριμένα, στο πρώτο μπλοκ έχουμε 32 φίλτρα, στο δεύτερο 64 και στο τρίτο 128. Το Padding χρησιμοποιείται στα συνεκτικά στρώματα για να διασφαλιστεί ότι το ύψος και το πλάτος των χαρτών χαρακτηριστικών εξόδου ταιριάζει με τις εισόδους. Η συνάρτηση ενεργοποίησης που χρησιμοποιείται σε όλα τα συνεκτικά επίπεδα είναι η ReLU. Κάθε μπλοκ ακολουθείται από ένα στρώμα Dropout, το οποίο σταδιακά κλιμακώνεται. Στο πρώτο έχουμε ποσοστό Dropout 0,2, στο δεύτερο 0,3, στο τρίτο 0,4. Αυτά τα μπλοκ ακολουθούνται από ένα πυκνό επίπεδο 128 κόμβων, ένα επίπεδο κανονικοποίησης, ένα στρώμα Dropout με ρυθμό ίσο με 0,5 και ένα πυκνό στρώμα με δύο κόμβους και συνάρτηση ενεργοποίησης Softmax. Επίσης, μετά από κάθε συνεκτικό στρώμα έχουμε ένα επίπεδο κανονικοποίησης δέσμης. Χρησιμοποιείται ο βελτιστοποιητής Adam με ρυθμό μάθησης 0,0005, η αρχικοποίηση πυρήνα he normal και η συνάρτηση απώλειας είναι η κατηγορική διασταυρούμενη εντροπία.



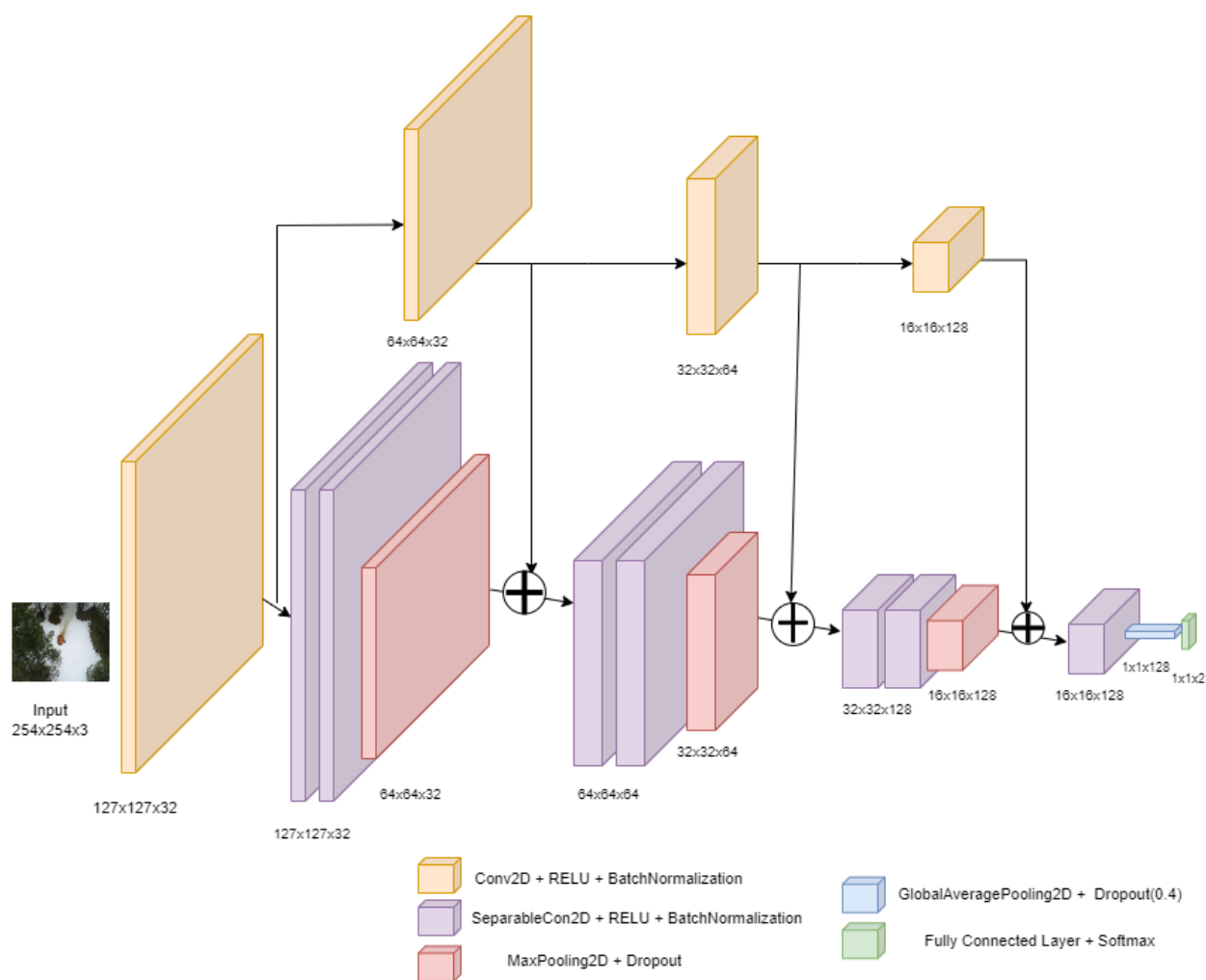
Εικόνα 5.7: Διάγραμμα του τρίτου μοντέλου για την πιστότητα και την απώλεια στα σύνολα εκπαίδευσης και επικύρωσης ως προς τις εποχές



Εικόνα 5.8: Πίνακας Σύγχυσης του τρίτου μοντέλου

Η μορφή αυτού του μοντέλου προέκυψε σταδιακά. Στην αρχή δοκιμάστηκαν λιγότερα μπλοκ και με λιγότερα φίλτρα, χωρίς padding, χωρίς κανονικοποίηση δέσμης, χωρίς επίπεδο εγκατάλειψης και χωρίς προεπεξεργασία δεδομένων, που στην περίπτωση μας, είναι το “Scaling” που είναι η κλιμάκωση από έναν αριθμό μεταξύ μηδέν και ένα. Έγιναν διάφορες δοκιμές με το πλήθος των φίλτρων, με καλύτερη απόδοση αυτού που επιλέχθηκε (32 , 64 , 128). Με την προεξεργασία είχαμε μια βελτίωση στην ακρίβεια και αισθητή μείωση στο σφάλμα. Με προσθήκη και dropout σταθερό στο 0,2 είχαμε μια ελάχιστη βελτίωση στην ακρίβεια και με κλιμακωτή αύξηση στο dropout (0,2, 0,3, 0,4, 0,5) μια ακόμα καλύτερη βελτίωση. Με επιπλέον κάποιες δοκιμές στο πλήθος των πλήρως διασυνδεδεμένων επιπέδων, του βελτιστοποιητή, του ρυθμού εκμάθησης και του αρχικοποιητή πυρήνα καταλήξαμε στο παρακάτω βέλτιστο αποτέλεσμα για αυτήν την αρχιτεκτονική.

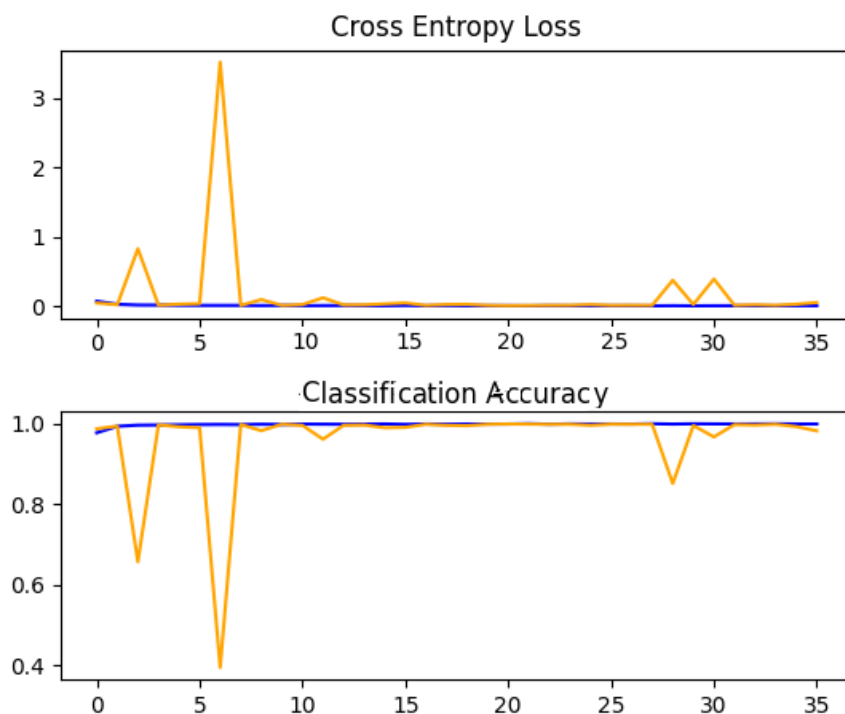
Στην Εικόνα 5.7 φαίνεται το διάγραμμα της απώλειας για τα δεδομένα εκπαίδευσης με μπλε και τα δεδομένα επικύρωσης με πορτοκαλί καθώς και το διάγραμμα της πιστότητας για τα δεδομένα εκπαίδευσης και επικύρωσης σε συνάρτηση με τις εποχές. Η εκπαίδευση σταματά πριν προλάβουν τα δεδομένα επικύρωσης να αποκλίσουν από αυτά της εκπαίδευσης. Στην Εικόνα 5.8 έχουμε τον πίνακα σύγκρισης του τρίτου μοντέλου απ’τον οποίο μπορούμε να υπολογίσουμε τις μετρικές που παρουσιάζονται συγκεντρωτικά στον Πίνακα 5.3.



Εικόνα 5.9: Τέταρτο μοντέλο

Το τέταρτο μοντέλο, η αρχιτεκτονική του οποίου απεικονίζεται στην Εικόνα 5.9, προήλθε από τη μελέτη του δικτύου Xception. Η τιμή των RGB εικόνων εισόδου, σε διαφορετικά κανάλια, κλιμακώνεται σε έναν αριθμό μεταξύ 0 και 1. Τα κρυφά επίπεδα βασίζονται σε διαχωρίσιμες ανα κανάλι συνελίξεις κατά βάθος και σε συντόμευση μεταξύ των μπλοκ συνελίξης.

Πιο αναλυτικά, μετά την κλιμάκωση περιλαμβάνει ένα δισδιάστατο συνελκτικό επίπεδο με μέγεθος 32 και βήμα 2. Στη συνέχεια, ακολουθούν επαναλαμβανόμενα μπλοκ που περιλαμβάνουν 2 διαχωρίσιμα κατά βάθος συνελκτικά επίπεδα με μικρά φίλτρα  $3 \times 3$ , όπου το καθένα ακολουθείται από κανονικοποίηση δέσμης και ενεργοποίηση ReLU. Η κανονικοποίηση δέσμης χρησιμοποιείται για να επιταχύνει τη διαδικασία εκπαίδευσης και για να φέρει περισσότερη τυχαιότητα, μειώνοντας τη σημασία των αρχικών βαρών και ρυθμίζοντας το μοντέλο. Ύστερα, στο κάθε μπλοκ υπάρχει ένα στρώμα υποδειγματοληψίας μεγίστου. Το πλήθος αυτών των μπλοκ είναι 3 και ο αριθμός των φίλτρων σε κάθε μπλοκ αυξάνεται με το βάθος του δικτύου. Πιο συγκεκριμένα, στο πρώτο μπλοκ έχουμε 32 φίλτρα, στο δεύτερο 64 και στο τρίτο 128. Αυτά τα μπλοκ συνδέονται με την πράξη της πρόσθεσης μέσω μιας σημειακής συνελίξης, μέσω μιας σύνδεσης καταλοίπου, με αρχή της σύνδεσης το προηγούμενο μπλοκ και τέλος το επόμενο μπλοκ. Μετά τα επαναλαμβανόμενα μπλοκ, έχουμε ένα στρώμα διαχωρίσιμης κατά βάθος συνελίξης με 128 φίλτρα, συνάρτηση ReLU, ακολουθούμενο από ένα επίπεδο κανονικοποίησης και ένα καθολικό επίπεδο υποδειγματοληψίας μέσου όρου.

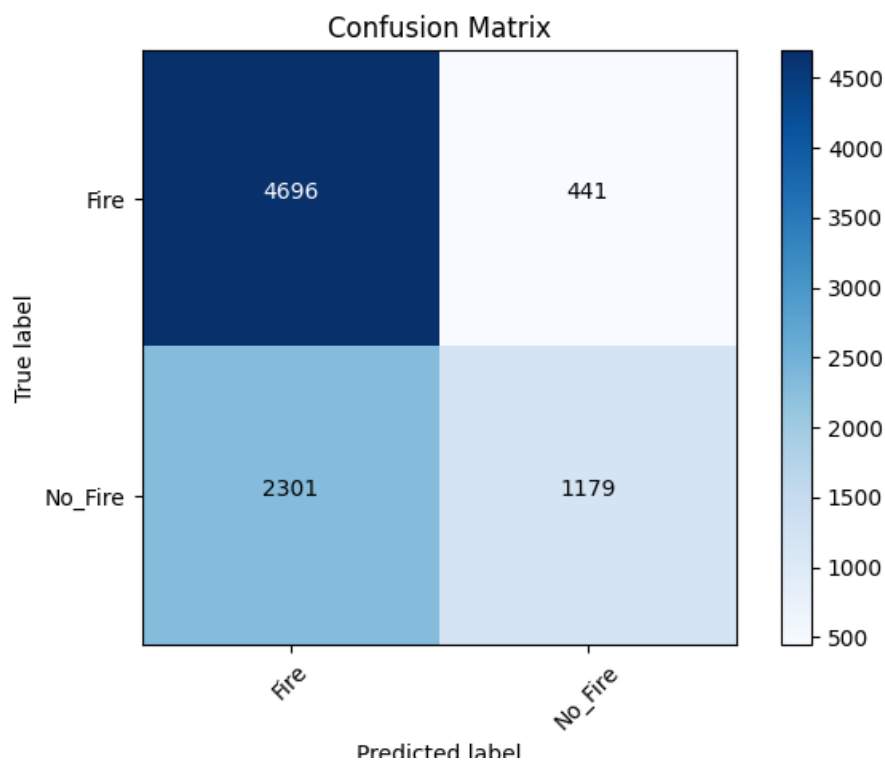


Εικόνα 5.10: Διάγραμμα του τέταρτου μοντέλου για την πιστότητα και την απώλεια στα σύνολα εκπαίδευσης και επικύρωσης ως προς τις εποχές

Τέλος, έχουμε ένα επίπεδο Dropout με ρυθμό 0,4 και ένα πλήρως διασυνδεδεμένο επίπεδο με δύο κόμβους και συνάρτηση ενεργοποίησης Softmax. Το Padding χρησιμοποιείται σε



όλα τα συνελκτικτικά στρώματα για να διασφαλιστεί ότι το ύψος και το πλάτος των χαρτών χαρακτηριστικών εξόδου ταιριάζει με τις εισόδους. Χρησιμοποιείται ο βελτιστοποιητής Nadam με ρυθμό μάθησης 0,001, η αρχικοποίηση πυρήνα he normal και η συνάρτηση απώλειας είναι η κατηγορική διασταυρούμενη εντροπία.



Εικόνα 5.11: Πίνακας Σύγχυσης του τέταρτου μοντέλου

Στην Εικόνα 5.10 παρατηρούμε, για το τέταρτο και τελευταίο μοντέλο, το διάγραμμα της απώλειας για τα δεδομένα εκπαίδευσης και επικύρωσης καθώς και το διάγραμμα της ακρίβειας για τα δεδομένα αυτά σε συνάρτηση με τις εποχές.

Ακόμη, παρατίθεται ο πίνακας σύγχυσης στην Εικόνα 5.11, στον οποίο είναι συγκεντρωμένες όλες οι τιμές TP, TN, FP, FN.

## 5.4 Συγκεντρωτικά Αποτελέσματα

Σε αυτή την ενότητα παρουσιάζονται συνοπτικά τα αποτελέσματα για τις υλοποιήσεις. Πιο συγκεκριμένα, στον Πίνακα 5.3 απεικονίζονται συγκεντρωτικά οι τρεις μετρικές αξιολόγησης -Πισιότητα, Ανάκληση-F1-score- οι οποίες προέκυψαν από τους Πίνακες σύγχυσης 5.2, 5.4, 5.11, 5.8 με τις Εξισώσεις 5.1-5.3 καθώς και οι παράμετροι όλων των καλύτερων διαμορφώσεων των μοντέλων που δοκιμάστηκαν κατά τη διάρκεια της πειραματικής διαδικασίας. Το 'Μοντέλο 1' είναι το μοντέλο Xception με μεταφορά μάθησης. Το 'Μοντέλο 2' είναι το μοντέλο MobileNetV2 με μεταφορά μάθησης. Το 'Μοντέλο 3' είναι το πρώτο μοντέλο που κατασκευάστηκε από την αρχή με τα απλά συνελκτικτικά επίπεδα. Το 'Μοντέλο 4' είναι το μοντέλο που κατασκευάστηκε από την αρχή με τις κατά βάθος διαχωρίσιμες συνελίξεις, εμπνευσμένο από την αρχιτεκτονική Xception. Από τα τέσσερα μοντελα, παρατηρούμε ότι το

βέλτιστο είναι το πρώτο μοντέλο, αφού έχει την καλύτερη πιστότητα, την καλύτερη ανάκληση αλλά και το υψηλότερο F1-score.

	<b>Πιστότητα</b>	<b>Ανάκληση</b>	<b>F1-score</b>	<b>Παράμετροι</b>
Μοντέλο 1	0,801	0,821	0,831	23.485.482
Μοντέλο 2	0,751	0,811	0,795	3.571.778
Μοντέλο 3	0,597	0,728	0,683	16.034.722
Μοντέλο 4	0,682	0,914	0,774	69.122

Πίνακας 5.3: Μειτρικές αξιολόγησης και παράμετροι όλων των μοντέλων

# Συμπεράσματα και Μελλοντικές Κατευθύνσεις

---

## 6.1 Συμπεράσματα

Αντικείμενο της παρούσας διπλωματικής εργασίας ήταν η κατασκευή και αξιολόγησης συστημάτων, τα οποία θα αναγνωρίζουν την ύπαρξη ή όχι φωτιάς σε εικόνες, οι οποίες έχουν ληφθεί από μεγάλο ύψος. Παράλληλα δόθηκε μια λεπτομερή ανάλυση των επιμέρους αρχιτεκτονικών τους συστατικών.

Ο σχεδιασμός του συστήματος οδήγησε στην παραγωγή μοντέλων με καλές επιδόσεις. Πιο συγκεκριμένα μελετήθηκαν δύο αρχιτεκτονικές συνελκτικών νευρωνικών δικτύων με μεταφορά μάθησης, τα Xception και MobileNetV2 όπου με αυτές επιτεύχθηκαν τα βέλτιστα αποτελέσματα. Επίσης, μελετήθηκαν και δυο νέες αρχιτεκτονικές κατασκευασμένες από την αρχή, χωρίς καμία προεκπαίδευση, με αποτελέσματα υστερούσαν ελαφρώς, αλλά ωστόσο και πάλι εμφάνιζαν πολύ ικανοποιητικά αποτελέσματα, με λιγότερες εκπαιδευσιμες παραμέτρους.

Το σύνολο δεδομένων που χρησιμοποιήθηκε περιλαμβάνει εικόνες οι οποίες έχουν τραβηχτεί αποκλειστικά σε ένα δασικό περιβάλλον, μια συγκεκριμένη χρονική περίοδο. Ακόμη, το σύνολο εκπαίδευσης περιέχει εικόνες επισημασμένες με 'Φωτιά' στις οποίες δεν φαίνεται ξεκάθαρα η φωτιά ούτε με το ανθρώπινο μάτι. Σε εικόνες είναι δύσκολη η διάκριση του καπνού σε σχέση με την ομίχλη και εύκολα μπορεί το σύστημα να 'μπερδευτεί' όταν δεν υπάρχει έντονα το χρώμα της φωτιάς ή όταν ο καπνός δεν είναι αρκετά σκουρόχρωμος. Αυτό δικαιολογεί και την πιθανή απόκλιση-σφάλμα στα δίκτυα που εκπαιδεύονται.

Υπήρξαν αρκετές περιπτώσεις λοιπόν που είχαμε λανθασμένες ταξινομήσεις των εικόνων για τους παραπάνω λόγους, έτσι μπορούμε να αποδώσουμε τις αστοχίες του συστήματος σε δύο κατευθύνσεις:

- Μη-αντιπροσωπευτικά δεδομένα: Τα δεδομένα περιλαμβάνουν εικόνες οι οποίες έχουν ανακτηθεί κατά τη διάρκεια προδιαγεγραμμένης καύσης πασσάλων, στη Βόρεια Αριζόνα των ΗΠΑ, δηλαδή οι εικόνες είναι μόνο από αυτό το δασικό περιβάλλον, μια συγκεκριμένη χρονική περίοδο. Συνεπώς, τα υπό εκπαίδευση μοντέλα δεν είχαν στη διάθεσή τους πολύ μεγάλη ποικιλία προτύπων εισόδου, ώστε να αποκτήσουν ένα γενικό χαρακτήρα. Για παράδειγμα, εικόνες που απεικονίζουν τον ήλιο, το ηλιοδασίλεμα και άλλα αντικείμενα παρόμοιου χρώματος με τη φλόγα, εικόνες φωτιάς σε κάποιο άλλο μέρος θα έπρεπε να υπάρχουν στο σύνολο δεδομένων ώστε να κατηγοριοποιηθούν στη σωστή κατηγορία.

- Περιορισμένοι υπολογιστικοί πόροι: Οι δοκιμές των διάφορων υλοποιήσεων έγιναν σε τοπικό υπολογιστή περιορισμένων δυνατοτήτων. Για τη διασφάλιση μιας καλύτερης υλοποίησης, ειδικά αν αυτό είχε ως στόχο εθνικά συμφέροντα, θα χρησιμοποιούνταν περισσότεροι υπολογιστικοί πόροι με δυνατότητα χρήσης περισσότερων GPU. Ειδικά για τις υλοποιήσεις των δικτύων από την αρχή, η μνήμη του υπολογιστή δεν επέτρεπε περισσότερους εις βάθος πειραματισμούς και πιο περίπλοκες αρχιτεκτονικές.

## 6.2 Μελλοντικές Επεκτάσεις

Στο πλαίσιο της μελλοντικής εργασίας πάνω στο πεδίο, διακρίνουμε τις ακόλουθες κατευθύνσεις ανάπτυξης, που μπορούν να συνεισφέρουν, μεμονωμένα ή συνδυαστικά, στην βελτίωση των παραγόμενων συστημάτων και την επέκταση των εφαρμογών τους:

- Προεπεξεργασία δεδομένων: Σε μεταγενέστερη εργασία θα μπορούσε να γίνει πληρέστερη προεπεξεργασία δεδομένων. Στην παρούσα εργασία, επειδή το σύνολο δεδομένων περιέχει περίπου 50 χιλιάδες εικόνες και η δοκιμή των μοντέλων σε αυτές είναι μια πολύ χρονοβόρα διαδικασία, δοκιμάστηκε μόνο το *scaling* ως μέθοδος προεπεξεργασίας. Άλλοι μέθοδοι που θα ήταν χρήσιμοι, είναι διάφορες τεχνικές επαύξησης δεδομένων.
- Υπερπαραμέτροι: Δοκιμές και άλλων συνδιασμών υπερπαραμέτρων θα μπορούσαν σίγουρα να γίνουν, όπως η περαιτέρω εξερεύνηση της υπερπαραμέτρου *alpha* του *MobileNetV2*.
- Χρήση διαφορετικών αρχιτεκτονικών: Η μηχανική μάθηση και η όραση υπολογιστών είναι δύο ραγδαία εξελισσόμενοι επιστημονικοί τομείς και γι' αυτό υπάρχουν και αναπτύσσονται πολλές αρχιτεκτονικές. Στα τελευταία στάδια της εργασίας, δημοσιεύτηκε πλήθος νέων διαθέσιμων μοντέλων από το *Keras*, για παράδειγμα διατέθηκε η 3η έκδοση της αρχιτεκτονικής *MobileNet*. Επομένως, σε μελλοντική εργασία θα μπορούσε να γίνει μελέτη και αξιολόγηση περαιτέρω αρχιτεκτονικών δικτύων, για να εντοπιστεί η κορυφαία διαθέσιμη μέθοδος.
- Συλλογή δεδομένων: Όπως είδαμε παραπάνω, είναι ιδιαίτερα σημαντικό να έχουμε διαθέσιμα δεδομένα που είναι όσο το δυνατόν πιο αντιπροσωπευτικά. Θα ήταν επωφελής η χρήση επιπλέον δεδομένων, εκτός του συνόλου δεδομένων που χρησιμοποιήθηκε, που να απεικονίζουν τη φωτιά και σε άλλα μέρη εκτός του δασικού περιβάλλοντος. Να υπάρχουν ακόμη εικόνες από νυχτερινές ώρες αλλά και εικόνες που δεν περιέχουν φωτιά ή καπνό αλλά είναι παραπλήσιες αυτής.

## Βιβλιογραφία

---

- [1] *Neuron of Human brain*. <https://el.m.wikipedia.org/wiki/%CE%9D%CE%B5%CF%85%CF%81%CF%8E%CE%BD%CE%B1%CF%82>. Ημερομηνία πρόσβασης: 09-06-2022.
- [2] *Artificial Neural Network*. <https://pl.m.wikipedia.org/wiki/Perceptron>. Ημερομηνία πρόσβασης: 09-06-2022.
- [3] *Underfitting and Overfitting*. <https://medium.com/ml-research-lab/under-fitting-over-fitting-and-its-solution-dc6191e34250>, note = Ημερομηνία πρόσβασης: 09-06-2022.
- [4] *Data Augmentation*. [https://www.researchgate.net/figure/Data-augmentation-using-semantic-preserving-transformation-for-SBIR\\_fig2\\_319413978](https://www.researchgate.net/figure/Data-augmentation-using-semantic-preserving-transformation-for-SBIR_fig2_319413978). Ημερομηνία πρόσβασης: 09-06-2022.
- [5] *Dropout*. <https://www.oreilly.com/library/view/tensorflow-for-deep/9781491980446/ch04.html>. Ημερομηνία πρόσβασης: 09-06-2022.
- [6] *Convolutional Neural Network architecture*. <https://www.javatpoint.com/pytorch-convolutional-neural-network>. Ημερομηνία πρόσβασης: 09-06-2022.
- [7] *Convolution 2 dimensions*. [https://www.researchgate.net/figure/Example-of-a-2-dimensional-convolution-The-filter-size-is-3x3-and-it-uses-a-0-padding\\_fig4\\_335259576](https://www.researchgate.net/figure/Example-of-a-2-dimensional-convolution-The-filter-size-is-3x3-and-it-uses-a-0-padding_fig4_335259576). Ημερομηνία πρόσβασης: 09-06-2022.
- [8] *Max and Average Pooling*. [https://www.researchgate.net/figure/Pooling-layer-operation-approaches-1-Pooling-layers-For-the-function-of-decreasing-the\\_fig4\\_340812216](https://www.researchgate.net/figure/Pooling-layer-operation-approaches-1-Pooling-layers-For-the-function-of-decreasing-the_fig4_340812216). Ημερομηνία πρόσβασης: 09-06-2022.
- [9] *Flatten Layer*. <https://medium.com/@cdabakoglu/what-is-convolutional-neural-network-cnn-with-keras-cab447ad204c>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [10] *LeNet Architecture*. <https://livebook.manning.com/book/grokking-deep-learning-for-computer-vision/chapter-5/v-3/22>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [11] *AlexNet Architecture*. <https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [12] *GoogleNet Architecture and Inception Module*. <https://ai.plainenglish.io/googlenet-inceptionv1-with-tensorflow-9e7f3a161e87>. Ημερομηνία Πρόσβασης: 09-06-2022.

- [13] *VGG Architecture*. <https://neurohive.io/en/popular-networks/vgg16/>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [14] *ResNet Architecture*. <https://medium.com/analytics-vidhya/deep-residual-learning-for-image-recognition-resnet-94a9c71334c9>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [15] <https://machinelearningmastery.com/transfer-learning-for-deep-learning/>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [16] *Depthwise Separable Convolution*. <https://medium.com/@zurister/depth-wise-convolution-and-depth-wise-separable-convolution-37346565d4ec>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [17] *Xception Architecture*. <https://towardsdatascience.com/xception-from-scratch-using-tensorflow-even-better-than-inception-940fb231ced9>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [18] *MobileNetV1 Architecture*. <https://blueardour.github.io/2019/04/17/mobilenet-v1.html>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [19] *Bottleneck Residual Block*. <https://paperswithcode.com/method/mobilenetv2>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [20] *Residual Block and Inverted Residual Block*. [https://www.researchgate.net/figure/Residual-block-12-36-and-inverted-residual-block\\_fig2\\_358518820](https://www.researchgate.net/figure/Residual-block-12-36-and-inverted-residual-block_fig2_358518820). Ημερομηνία Πρόσβασης: 09-06-2022.
- [21] *MobileNetV2 Architecture*. <https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [22] *THE FLAME DATASET: AERIAL IMAGERY PILE BURN DETECTION USING DRONES (UAVS)*. <https://ieee-dataport.org/open-access/flame-dataset-aerial-imagery-pile-burn-detection-using-drones-uavs>. Ημερομηνία πρόσβασης: 09-06-2022.
- [23] *Confusion Matrix*. <https://shravantandale456.medium.com/importance-of-confusion-matrix-in-machine-learning-and-cybersecurity-80e67f5858fb>. Ημερομηνία Πρόσβασης: 09-06-2022.
- [24] *National interagency fire center*. <https://www.nifc.gov/fireInfo/nfn.html>. Ημερομηνία πρόσβασης: 09-06-2022.
- [25] *National interagency fire center*. [https://www.nifc.gov/fireInfo/fireInfo\\_statistics.html](https://www.nifc.gov/fireInfo/fireInfo_statistics.html). Ημερομηνία πρόσβασης: 09-06-2022.
- [26] J. Toledo Castro P. Caballero Gil N. Rodríguez Pérez I. Santos González C. Hernández Goya R. Aguasca-Colomo. *Forest fire prevention, detection, and fighting based on fuzzy logic and wireless sensor networks*. *Hindawi Complexity*, σελίδες 1–17, 2018.

- [27] H. Kaur S.K. Sood. *Fog-assisted IoT-enabled scalable network infrastructure for wildfire surveillance*. *J. Netw. Comput. Appl.*, 144:171–183, 2019.
- [28] J.L. Coen W. Schroeder S.D. Rudlosky. *Transforming wildfire detection and prediction using new and underused sensor and data sources integrated with modeling in: Handbook of Dynamic Data Driven Applications Systems*. Springer, 9η έκδοση, 2018.
- [29] Q. Huang A. Razi F. Afghah P. Fule. *Wildfire spread modeling with aerial image processing*. *IEEE 19th European Signal Processing Conference*, Barcelona, Spain, 2011.
- [30] Y. H. Habiboglu O. Gunay και A. E. Cetin. *Real-time wildfire detection using correlation descriptors*. *IEEE 21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, Cork, Ireland, 2020.
- [31] D. Twidwell C. Allen και J. Higgins C. L. S. E. C. Detweiler. *Smokey comes of age: unmanned aerial systems for fire management*. *Frontiers in Ecology and the Environment*, 2016.
- [32] M. P'olka S. Ptak και Łukasz Kuziora. *The use of uav's for search and rescue operations*. *12th international scientific conference of young scientists on sustainable, modern and safe transport*, High Tatras, Slovakia, 2017.
- [33] G. Hristov J. Raychev D. Kinaneva και P. Zahariev. *Emerging methods for early detection of forest fires using unmanned aerial vehicles and lorawan sensor networks*. *28th EAEEIE Annual Conference (EAEEIE)*, Reykjavik, Ireland, 2018.
- [34] M. Erdelj E. Natalizio K. R. Chowdhury και I. F. Akyildiz. *Help from the sky: Leveraging uavs for disaster management*. *IEEE Pervasive Computing*, 16:24–32, 2017.
- [35] M. Erdelj και E. Natalizio. *Uav-assisted disaster management: Applications and open issues*. *International Conference on Computing, Networking and Communications (ICNC)*, Kauai, Hawaii, USA, 2016.
- [36] A. Shamsoshoara M. Khaledi F. Afghah A. Razi J. Ashdown. *Distributed cooperative spectrum sharing in uav networks using multi-agent reinforcement learning*. *16th IEEE Annual Consumer Communications and Networking Conference, CCNC, IEEE*, Las Vegas, NV, USA, 2019.
- [37] A. Shamsoshoara M. Khaledi F. Afghah A. Razi J. Ashdown K. Turck. *A solution for dynamic spectrum management in mission-critical uav networks*. *16th Annual IEEE International Conference on Sensing, Communication, and Networking, SECON, IEEE*, Boston, MA, USA, 2019.
- [38] A. Shamsoshoara F. Afghah A. Razi S. Mousavi J. Ashdown K. Turk. *An autonomous spectrum management scheme for unmanned aerial vehicle networks in disaster relief operations*. *IEEE*, 2020.

- [39] F. Afghah A. Razi J. Chakareski και J. Ashdown. *Wildfire monitoring in remote areas using autonomous unmanned aerial vehicles*. *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, Paris, France, 2019.
- [40] S. Islam και A. Razi. *A path planning algorithm for collective monitoring using autonomous drones*. *53rd Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, USA, 2019.
- [41] S. Islam Q. Huang F. Afghah P. Fule και A. Razi. *Fire frontline monitoring by enabling uav-based virtual reality with adaptive imaging rate*. *Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2019.
- [42] A. Shamsoshoara M. Khaledi F. Afghah A. Razi J. Ashdown και K. Turck. *A solution for dynamic spectrum management in missioncritical uav networks*. *IEEE*, 2019.
- [43] Phil Simon. *Too Big to Ignore: The Business Case for Big Data*. Wiley, 2013.
- [44] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.
- [45] Mehryar Mohri Afshin Rostamizadeh Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, 2018.
- [46] S. Haykin. *Νευρωνικά Δίκτυα και Μηχανική Μάθηση*. Παπασωτηρίου, 3η έκδοση, 2010.
- [47] Rosenblatt Frank. *The Perceptron—a perceiving and recognizing automaton*. *Cornell Aeronautical Laboratory*, 1957.
- [48] Διαμαντάρας, Κ. *Τεχνητά Νευρωνικά Δίκτυα*. Κλειδάριθμος, 2007.
- [49] Ματσοατίνης Ν. *Συστήματα Υποστήριξης Αποφάσεων*. Εκδόσεις Νέων Τεχνολογιών, 2010.
- [50] Risto Miikkulainen, Jason Liang, Elliot Meyerson, Aditya Rawal, Dan Fink, Olivier Francon, Bala Raju, Hormoz Shahrzad, Arshak Navruzyan, Nigel Duffy και Babak Hodjat. *Evolving Deep Neural Networks*. Amsterdam: Elsevier, 2018.
- [51] Nitish Srivastava Geoffrey Hinton Alex Krizhevsky Ilya Sutskever Ruslan Salakhutdinov. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. *Journal of Machine Learning Research*, 2014.
- [52] Chervyakov N.I. Valueva M.V. Nagornov N.N. Lyakhov P.A., Valuev G.V. *Application of the residue number system to reduce hardware costs of the convolutional neural network implementation*. Elsevier, 2020.
- [53] Aston Zhang Zack C. Lipton Mu Li Alex J. Smola. *Dive into Deep Learning*. Berkeley, 2019.



- [54] Kunihiko Fukushima. *Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*. Springer, 1980.
- [55] *THE MNIST DATABASE*. <https://yann.lecun.com/exdb/mnist/>. Ημερομηνία πρόσβασης: 09-06-2022.
- [56] Yann LeCun Leon Bottou Yoshua Bengio και Patrick Haffner. *Gradient-Based Learning Applied to Document Recognition*. IEEE, 1998.
- [57] Schmidhuber Jürgen. *A Fast Learning Algorithm for Deep Belief Nets*. MIT Press Direct, 2015.
- [58] Oh KS Jung K. *GPU implementation of neural networks*. Elsevier, 2004.
- [59] Kumar Chellapilla Sid Puri Patrice Simard. *High Performance Convolutional Neural Networks for Document Processing*. HAL Open Science, 2006.
- [60] Alex Krizhevsky Ilya Sutskever Geoffrey E. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. 26th Conference on Neural Information Processing Systems, Harrahs and Harveys, Lake Tahoe, Nevada, USA, 2012.
- [61] *ImageNet*. <https://www.image-net.org>. Ημερομηνία πρόσβασης: 09-06-2022.
- [62] Christian Szegedy Wei Liu Yangqing Jia Pierre Sermanet Scott Reed Dragomir Anguelov Dumitru Erhan Vincent Vanhoucke Andrew Rabinovich. *Going Deeper with Convolutions*. 28th Conference on Neural Information Processing Systems, Palais des Congrès de Montréal, Montréal CANADA, 2014.
- [63] Karen Simonyan και Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 28th Conference on Neural Information Processing Systems, Palais des Congrès de Montréal, Montréal CANADA, 2014.
- [64] Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun. *Deep Residual Learning for Image Recognition*. 29th Conference on Neural Information Processing Systems, Palais des Congrès de Montréal, Montréal CANADA, 2014.
- [65] G. Healey D. Slater T. Lin B. Drda και A.D. Goedeke. *A system for real-time fire detection*. *Proceedings of the International Conference Computer Vision and Pattern Recognition*, Berlin, Germany, 1993.
- [66] T Chen P. Wu και Y. Chiou. *An early fire-detection method based on image processing*. *Proceedings of the International Conference Computer Vision and Pattern Recognition*, Berlin, Germany, 2004.
- [67] T Celik και H Demirel. *Fire detection in video sequences using a generic color model*. *Fire Safety Journal*, 2009.
- [68] M.M. Umar L.C.D. Silva M.S.A. Bakar M.I. Petra. *State of the art of smoke and fire detection using image processing*. *International Journal of Signal and Imaging Systems Engineering*, 2017.

- [69] N.I.binti Zaidi N.A.A. binti Lokman M.R. bin Daud H. Achmad K.A. Chia. *Fire recognition using RGB and YCbCr color space*. *ARPN Journal of Engineering and Applied Sciences (JEAS)*, 2015.
- [70] F. Yuan. *An integrated fire detection and suppression system based on widely available video surveillance*. *Machine Vision and Applications*, 2010.
- [71] J. R. Martinezde Dios B. C. Arrue A. Ollero L. Merino και F. Gómez-Rodríguez. *Computer vision techniques for forest fire perception*. *Image and Vision Computing*, 2008.
- [72] T Celik H Demirel H Ozkaramanli και M Uyguroglu. *Fire detection using statistical color model in video sequences*. *Journal of Visual Communication and Image Representation*, 2007.
- [73] C. Yuan Z. Liu Y. Zhang. *UAV-Based forest fire detection and tracking using image processing techniques*. *International Conference on Unmanned Aircraft Systems (ICUAS), IEEE, Denver, Colorado, USA, 2015*.
- [74] W. Phillips M. Shah και N.da Vitoria Lobo. *Flame recognition in video*. *Pattern Recognition Letters, Elsevier, 2002*.
- [75] C. Liu και N. Ahuja. *Vision based fire detection*. *Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 2004*.
- [76] B.U. Toreyin Y. Dedeoglu και A.E. Cetin. *Flame detection in video using hidden Markov models*. *Proceedings of the International Conference on Image Analysis and Processing (ICIAP), Cagliari, Italy, 2005*.
- [77] G Marbach M Loepfe και T Brupbacher. *An image processing technique for fire detection in video images*. *Fire Safety Journal, 2006*.
- [78] B. Toreyin Y. Dedeoglu U. Gudukbay και A. Cetin. *Computer vision based method for real-time fire and flame detection*. *Pattern Recognition Letters, 2006*.
- [79] Y. H. Habiboglu O. Gunay και A. E. Cetin. *Real-time wildfire detection using correlation descriptors*. *19th European Signal Processing Conference, IEEE, Barcelona, Spain, 2011*.
- [80] B. Ko K. Cheong και J. Nam. *Fire detection based on vision sensor and support vector machines*. *Fire Safety Journal, 2009*.
- [81] D. Zhang S. Han J. Zhao Z. Zhang C. Qu Y. Ke και X. Chen. *Image based forest fire detection using dynamic characteristics with artificial neural networks*. *International Joint Conference on Artificial Intelligence, Hainan, China, 2009*.
- [82] A. Chenebert T.P. Breckon και A. Gaszczak. *A non-temporal texture driven approach to real-time fire detection*. *18th IEEE International Conference on Image Processing, Brussels, Belgium, 2011*.

- [83] J. Choi και J.Y. Choi. *Patch-based fire detection with online outlier learning*. 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Karlsruhe, Germany, 2015.
- [84] Tom Toulouse Lucile Rossi Turgay Celik και Moulay Akhloufi. *Automatic fire pixel detection using image processing: a comparative analysis of rule-based and machine learning-based methods*. Signal, Image and Video Processing, Springer, 2016.
- [85] Sebastien Frizzi Rabeb Kaabi Moez Bouchouicha Jean Marc Ginoux Eric Moreau και Farhat Fnaiech. *Convolutional neural network for video fire and smoke detection*. 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON, IEEE), Florence, Italy, 2016.
- [86] Y. Luo L. Zhao P. Liu και D. Huang. *Fire smoke detection algorithm based on motion characteristic and convolutional neural networks*. Multimedia Tools and Applications, 2017.
- [87] Qingjie Zhang Jiaolong Xu Liang Xu και Haifeng Guo. *Deep convolutional neural networks for forest fire detection*. International Forum on Management, Education and Information Technology Application, Guangzhou, China, 2016.
- [88] Jivitesh Sharma Ole Christoffer Granmo Morten Goodwin και Jahn Thomas Fidje. *Deep convolutional neural networks for fire detection in images*. International Conference on Engineering Applications of Neural Networks (EANN), Athens, Greece, 2017.
- [89] Andrew J. Dunningts και Toby P. Breckon. *Experimentally Defined Convolutional Neural Network Architecture Variants for Non-Temporal Real-Time Fire Detection*. 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018.
- [90] Ganesh Samarth C.A. Neelanjan Bhowmik και Toby P. Breckon. *Experimental Exploration of Compact Convolutional Neural Network Architectures for Non-Temporal Real-Time Fire Detection*. 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, Florida, USA, 2019.
- [91] Abdulaziz Namozov και Young Im Cho. *An Efficient Deep Learning Algorithm for Fire and Smoke Detection with Limited Data*. Advances in Electrical and Computer Engineering (AECE), 2018.
- [92] W. Thomson Neelanjan Bhowmik και T. Breckon. *Efficient and Compact Convolutional Neural Network Architectures for Non-temporal Real-time Fire Detection*. 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 2020.
- [93] Haiyu Wu Huayu Li Alireza Shamsoshoara Abolfazl Razi και Fatemeh Afghah. *Transfer Learning for Wildfire Identification in UAV Imagery*. 54th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 2020.

- [94] Alireza Shamsoshoara Fatemeh Afghah Abolfazl Razi Liming Zheng Peter Z.Fulé και Erik Blasch. *Aerial imagery pile burn detection using deep learning: The FLAME dataset*. *Computer Networks*, Elsevier, 2021.
- [95] Francois Chollet. *Xception: Deep Learning with Depthwise Separable Convolutions*. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017.
- [96] Andrew G. Howard Menglong Zhu Bo Chen Dmitry Kalenichenko Weijun Wang Tobias Weyand Marco Andreetto και Hartwig Adam. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. -, 2017.
- [97] Diederik Kingma Jimmy Ba. *Adam: A Method for Stochastic Optimization*. *3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, 2015.
- [98] Timothy Dozat. *INCORPORATING NESTEROV MOMENTUM INTO ADAM*. *4th International Conference on Learning Representations (ICLR)*, San Juan, Puerto Rico, 2016.
- [99] Martín Abadi Paul Barham Jianmin Chenet et al. Google Brain. *TensorFlow: A System for Large-Scale Machine Learning*. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, Savannah, GA, USA, 2016.
- [100] Sharan Chetlur Cliff Woolley Philippe Vandermersch Jonathan Cohen John Tran Bryan Catanzaro Evan Shelhamer. *cuDNN: Efficient Primitives for Deep Learning*. *Deep Learning and Representation Learning Workshop (NIPS 2014)*, Montreal, CANADA, 2014.
- [101] Francois Chollet. *Introducing Keras: deep learning with Python*. Manning, 2017.
- [102] James Bergstra Olivier Breuleux Frédéric Bastien Pascal Lamblin Razvan Pascanu Guillaume Desjardins Joseph Turian David Warde Farley Yoshua Bengio. *Theano: A CPU and GPU Math Compiler in Python*. *9th PYTHON IN SCIENCE CONF. (SCIPY 2010)*, Austin, Texas., 2010.
- [103] Frank Seide Amit Agarwal. *CNTK: Microsoft's Open-Source Deep-Learning Toolkit*. *22nd ACM SIGKDD International Conference*, San Francisco, California, USA, 2016.
- [104] Fabian Pedregosa Gael Varoquaux Alexandre Gramfort Vincent Michel David Cournapeauet et al. *Scikit-learn: Machine Learning in Python*. *Journal of Machine Learning Research* 12, 2011.
- [105] Charles R Harris K. Jarrod Millmanet et al. *Array programming with NumPy*. *Nature*, 2020.
- [106] SciPy community. *SciPy Reference Guide*. -, 2012.

- [107] John D. Hunter. *Matplotlib: A 2D Graphics Environment*. IEEE, *Computing in Science and Engineering*, 2007.



## Απόδοση ξενόγλωσσων όρων

---

### Απόδοση

ακρίβεια  
ανάκληση  
ανεστραμμένο μπλοκ με κατάλοιπο  
από την αρχή  
αρχικοποιητής πυρήνα  
βελτιστοποιητής  
βήμα  
γέμισμα περιθωρίου  
εξαγωγή χαρακτηριστικών  
εξαφανιζόμενη κλίση  
εναέρια μη επανδρωμένα οχήματα  
ενισχυτική μάθηση  
επιβλεπόμενη μάθηση  
επίπεδο γραμμικής συμφόρησης  
επίπεδο εγκατάλειψης  
επίπεδο ενεργοποίησης  
επίπεδο κανονικοποίησης παρτίδας  
επίπεδο υποδειγματοληψίας μεγίστου  
εποχές  
κατά βάθος διαχωρίσιμη συνέλιξη  
κατηγορική διασταυρούμενη εντροπία  
μέγεθος παρτίδας  
μείωση διαστασιμότητας  
μεταφορά μάθησης  
μη επιβλεπόμενη μάθηση  
μηχανική μάθηση  
μπλοκ με κατάλοιπο  
παλινδρόμηση  
πίνακας σύγχυσης  
πιστότητα  
πλήρως συνδεδεμένο επίπεδο  
πολλαπλασιαστής βάθους  
πρόωρη διακοπή  
πυκνό επίπεδο

### Ξενόγλωσσος όρος

precision  
recall  
inverted residual block  
from scratch  
kernel initializer  
optimizer  
stride  
padding  
feature learning  
vanishing gradient  
unmanned aerial vehicles  
reinforcement learning  
supervised learning  
linear bottleneck layer  
dropout layer  
activation layer  
batch normalization layer  
max pooling layer  
epochs  
depth-wise separable convolution  
categorical cross entropy loss  
batch size  
dimensionality reduction  
transfer learning  
unsupervised learning  
machine learning  
residual block  
regression  
confusion matrix  
accuracy  
fully connected layer  
width multiplier  
early stopping  
dense layer

πυρήνας/φίλτρο	kernel/filter
ρυθμός μάθησης	learning rate
σημειακή συνέλιξη	pointwise convolution
συνάρτηση ενεργοποίησης	activation function
συνελικτικά νευρωνικά δίκτυα	convolutional neural networks
συνελικτικό επίπεδο	convolutional layer
συνέλιξη κατά βάθος	depthwise convolution
σύνολο εκπαίδευσης	train set
σύνολο ελέγχου	test set
σύνολο επικύρωσης	valid set
συσταδοποίηση	clustering
ταξινόμηση	classification
τυπική απόκλιση	standard deviation
υπερπροσαρμογή	overfitting
υποδειγματοληψία μέσου όρου	average pooling
υποπροσαρμογή	underfitting
χάρτης ενεργοποίησης/χαρακτηριστικών	activation/feature map