Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

# Ανάλυση Συναισθημάτων Σε Εικόνες Με Χρήση Γεννητικών Ανταγωνιστικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΕΛΕΝΑ Γ. ΒΕΡΓΟΠΟΥΛΟΥ**

**Επιβλέπων :** Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2021

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

# Ανάλυση Συναισθημάτων Σε Εικόνες Με Χρήση Γεννητικών Ανταγωνιστικών Δικτύων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΕΛΕΝΑ Γ. ΒΕΡΓΟΠΟΥΛΟΥ**

**Επιβλέπων :** Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 15η Νοεμβρίου.

| | | |
|---|---|---|
| Στέφανος Κόλλιας | Ανδρέας Γεώργιος Σταφυλοπάτης | Γεώργιος Στάμου |
| Καθηγητής Ε.Μ.Π. | Καθηγητής Ε.Μ.Π. | Καθηγητής Ε.Μ.Π. |

Αθήνα, Νοέμβριος 2021

**Έλενα Γ. Βεργοπούλου**
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

# Περίληψη

Η παρούσα διπλωματική πραγματεύεται το θέμα της ανάλυσης αλλά και σύνθεσης συναισθημάτων στο ανθρώπινο πρόσωπο με τη χρήση βαθιών νευρωνικών δικτύων. Ο βασικός σκοπός είναι η μετατροπή ενός συναισθήματος που αναπαριστάται σε μία εικόνα σε ένα άλλο νέο συναίσθημα, πετυχαίνοντας όμως παράλληλα ένα ρεαλιστικό αποτέλεσμα. Για την κατηγοριοποίηση των συναισθημάτων επιλέχθηκε η δομή των επτά βασικών εκφράσεων (Seven Basic Expressions). Η διαδικασία αυτή της μετατροπής ονομάζεται Μετάφραση Εικόνας-σε-Εικόνα (Image-to-Image Translation).

Τα τελευταία χρόνια σημαντικό ρόλο σε αυτή τη διαδικασία διαδραματίζουν τα Ανταγωνιστικά Γεννητικά Δίκτυα (Generative Adversarial Networks). Τα συστήματα αυτά χρησιμοποιούνται όλο και περισσότερο σε σημαντικές εφαρμογές και διακρίνονται για την ικανότητά τους να παραγάγουν υψηλής ποιότητας εικόνες. Αποτελούνται από δύο μοντέλα, έναν γεννήτορα (generator), ο οποίος είναι υπεύθυνος για το κομμάτι της σύνθεσης των εικόνων και έναν διευκρινιστή (discriminator), ο οποίος κάνει την κατηγοριοποίηση.

Για το πειραματικό κομμάτι χρησιμοποιήθηκε το μοντέλο StarGAN, το οποίο έχει τη δυνατότητα να εκτελεί μεταφράσεις εικόνας για πολλαπλούς τομείς χρησιμοποιώντας μόνο ένα μοντέλο, καθώς και το σύνολο δεδομένων AffectNet, μία μεγάλη βάση με πολλές ταυτότητες και εύρος, γεγονός που θα βοηθήσει το GAN στην εξαγωγή όσο το δυνατόν καλύτερων αποτελεσμάτων.

Εξετάζουμε την απόδοση του διευκρινιστή τόσο για αληθινές εικόνες όσο και για εικόνες που έχουν δημιουργηθεί από τον γεννήτορα. Τα αποτελέσματα που λαμβάνουμε είναι πολύ ενθαρρυντικά καθώς οι εικόνες που συνθέτουμε φαίνεται να είναι αρκετά ρεαλιστικές ώστε να καταφέρουν να εξαπατήσουν τον διευκρινιστή στον μεγαλύτερο βαθμό. Επιπλέον και ως προς την παραγωγή των ζητούμενων συναισθημάτων έχουμε θετικά αποτελέσματα και ο διευκρινιστής αναγνωρίζει επιτυχώς το εκάστοτε συναίσθημα. Όσον αφορά τον γεννήτορα πραγματοποιήθηκε μία ανάλυση ως προς το πόσο ρεαλιστικές είναι οι εικόνες και τα συναισθήματα που παράγει αλλά και κατά πόσο αλλοιώνει την αρχική εικόνα. Τα αποτελέσματα αυτά παρουσιάζονται αναλυτικά και παραθέτονται προτάσεις για μελλοντικές επεκτάσεις.

## Λέξεις κλειδιά

Τεχνητή Νοημοσύνη, Βαθιά Μάθηση, Γεννητικά Ανταγωνιστικά Δίκτυα, Μετάφραση Εικόνας-σε-Εικόνα, Αναγνώριση Συναισθημάτων.

# Abstract

This diploma thesis deals with the analysis and synthesis of emotions in the human face, using deep neural networks. The basic aim is to convert an emotion which is represented in one picture into a new emotion, while also achieving a realistic result. For the task of the emotion classification we chose the seven basic expression classification. The process of the conversion is called Image-to-Image Translation.

In the last couple of years, an important role in this process is played by Generative Adversarial Networks (GANS). These systems are increasingly used in important applications and are known for their ability to produce high-quality images. They consist of two models, a generator, which is responsible for the bit of image composition, and a discriminator, which does the classification.

For the experimental part we used the StarGAN model, which has the ability to perform image translations for multiple domains using only one model, as well as the dataset AffectNet, a large base with many identities and range, which will help our GAN to export as many results as possible.

We examine the discriminator's performance for both the true images and the images created by the generator. The results we are receiving are very encouraging as the images we compose appear to be realistic enough to cheat the discriminator to the greatest extent. In addition, we receive positive results on the production of the requested emotions as well since the discriminator successfully recognizes the respective emotions. As far as the generator is concerned, an analysis has been made of how realistic the images and the emotions it produces are, and whether it distorts the original picture. These results are presented in detail along with proposals for future extensions.

# Key words

Artificial Intelligence, Deep Learning,Generative Adversarial Networks,Image-to-image Translation, Emotion Recognition.

# Ευχαριστίες

*Με την παρούσα διπλωματική ολοκληρώνεται ένας πολύ σημαντικός κύκλος της ζωής μου, αυτός των προπτυχιακών μου σπουδών. Συνεπώς θα ήθελα να ευχαριστήσω όλα τα πρόσωπα που με υποστήριξαν τόσο στην εκπόνηση της διπλωματικής όσο και όλο αυτό το διάστημα.*

*Αρχικά θα ήθελα να ευχαριστήσω τον κ. Στέφανο Κόλλια, επιβλέποντα καθηγητή μου για την εμπιστοσύνη και την καθοδήγησή του. Παράλληλα χρωστάω ένα εξίσου μεγάλο ευχαριστώ στον διδάκτορα κ. Δημήτριο Κόλλια, ο οποίος υπήρξε σημαντική βοήθεια και με βοήθησε να ξεπεράσω όλα τα σημαντικά εμπόδια που συνάντησα. Επιπλέον, θα ήθελα να ευχαριστήσω τους καθηγητές, κ. Ανδρέα Σταφυλοπάτη και κ. Γεώργιο Στάμου για την παρουσία τους στην τριμελή επιτροπή εξέτασης.*

*Τέλος ευχαριστώ τόσο την οικογένεια μου όσο και τους φίλους μου για την συμπαράσταση, την υπομονή και την αγάπη τους.*

# Περιεχόμενα

**Περίληψη** **5**

**Abstract** **7**

**Ευχαριστίες** **9**

**Περιεχόμενα** **11**

**Κατάλογος Σχημάτων** **13**

**Κεφάλαιο 1. Εισαγωγή** **17**
1.1 Αντικείμενο Εργασίας 17
1.2 Αναδρομή 17
1.3 Δομή Εργασίας 18

**Κεφάλαιο 2. Δομές Συναισθημάτων** **21**
2.1 Εισαγωγή 21
2.2 Αναδρομή 21
2.3 Επτά Βασικά Συναισθήματα (Seven Basic Expressions) 22
2.3.1 Κατηγοριοποίηση 22
2.3.2 Ανάλυση 22
2.4 Μονάδες Δράσης Προσώπου (Facial Action Units) 26
2.4.1 Κατηγοριοποίηση 26
2.4.2 Ανάλυση 26
2.5 Σθένος και διέγερση (Valence and arousal) 28
2.5.1 Κατηγοριοποίηση 28
2.5.2 Ανάλυση 28
2.6 Συμπεράσματα 29

**Κεφάλαιο 3. Μετάφραση Εικόνας-σε-Εικόνα με χρήση Ανταγωνιστικών Γεννητικών Δικτύων** **31**
3.1 Εισαγωγή 31
3.2 Βαθιά Νευρωνικά Δίκτυα 31
3.2.1 Βαθιά Μάθηση (Deep Learning) 31
3.2.2 Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks) 31
3.3 Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks) 32
3.3.1 Εισαγωγή 32
3.3.2 Επιβλεπόμενη και μη-επιβλεπόμενη εκμάθηση 33
3.3.3 Διευκρινιστής και γεννήτορας 35
3.3.4 Εκπαίδευση (Training) 36
3.3.5 Συναρτήσεις απωλειών (Loss functions) 38
3.4 Μετάφραση Εικόνας-σε-Εικόνα (Image-to-Image Translation) 40

3.4.1 Ανάλυση 40
3.4.2 Μετάφραση με ζεύγη: pix2pix 41
3.4.3 Μετάφραση χωρίς ζεύγη: CycleGAN 43
3.4.5 Μετάφραση εικόνων με πολλαπλούς τομείς: StarGAN 45
3.5 Συμπεράσματα 50

**Κεφάλαιο 4. Σύνολα Δεδομένων 51**
4.1 Εισαγωγή 51
4.2 RAF-DB 51
4.3 AffectNet 52
4.4 AFF-WILD2 54
4.5 Συμπεράσματα 55

**Κεφάλαιο 5. Πειραματική Διαδικασία 57**
5.1 Εισαγωγή 57
5.2 Παράμετροι και υπερ-παράμετροι 57
5.2.1 Εισαγωγή 57
5.2.2 Εποχές, επαναλήψεις και μέγεθος παρτίδας 57
5.2.3 Ρυθμός εκμάθησης 59
5.2.4 Τύπος GAN και ποινή βαθμίδας 60
5.2.5 Χαρακτηριστικά και εικόνες 60
5.2.6 Επαύξηση δεδομένων 61
5.3 Στάδιο εκπαίδευσης 63
5.3.1 Ανάλυση 63
5.3.2 Εφαρμογή 63
5.4 Στάδιο δοκιμής 64
5.4.1 Ανάλυση 64
5.4.2 Εφαρμογή 64

**Κεφάλαιο 6. Αποτελέσματα και Αξιολόγηση 67**
6.1 Εισαγωγή 67
6.2 Αξιολόγηση διευκρινιστή με εικόνες από το σύνολο επικύρωσης 67
6.3 Αξιολόγηση διευκρινιστή με εικόνες από τον γεννήτορα 70
6.4 Αξιολόγηση γεννήτορα 72
6.4.1 Αξιολόγηση εικόνων 72
6.4.2 Μέσο τετραγωνικο σφάλμα 75

**Κεφάλαιο 7. Επίλογος 79**
7.1 Γενικά συμπεράσματα 79
7.2 Μελλοντικές Επεκτάσεις 79

**Βιβλιογραφία 81**

# Κατάλογος Σχημάτων

2.1: Χαρακτηριστικά του θυμού βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

2.2: Χαρακτηριστικά του φόβου βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

2.3: Χαρακτηριστικά της αηδίας βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

2.4: Χαρακτηριστικά της περιφρόνησης βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

2.5: Χαρακτηριστικά της χαράς βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

2.6: Χαρακτηριστικά της θλίψης βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

2.7: Χαρακτηριστικά της έκπληξης βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

2.8: Παράδειγμα αντιστοίχισης των AU με τις δράσεις τους (**Πηγή**: https://arxiv.org/pdf/2001.11409.pdf )

2.9: Τα επτά βασικά συναισθήματα με σύνθεση μονάδων δράσης (**Πηγή**: Wikipedia)

2.10: Η δισδιάστατη δομή των συναισθημάτων σύμφωνα με τον Russell (**Πηγή**: https://arxiv.org/ftp/arxiv/papers/2001/2001.04509.pdf)

3.1: Παράδειγμα επιβλεπόμενης μάθησης (**Πηγή**: machinelearningmastery)

3.2: Παράδειγμα μη επιβλεπόμενης μάθησης (**Πηγή**: machinelearningmastery)

3.3: Δομή ενός GAN (**Πηγή**: developers.google)

3.4: Παράδειγμα μετάφρασης εικόνας-σε-εικόνα (**Πηγή**: https://junyanz.github.io/CycleGAN/)

3.5: Μοντέλο pix2pix (**Πηγή**: Image-to-Image Translation with Conditional Adversarial Networks)

3.6: Αποτελέσματα pix2pix (**Πηγή**: Image-to-Image Translation with Conditional Adversarial Networks)

3.7: Παράδειγμα μεταφράσεων με και χωρίς ζεύγη (**Πηγή**: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks)

3.8: Συναρτήσεις και απώλειες του CycleGAN (**Πηγή**: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks)

3.9: Παράδειγμα μετάφρασης εικόνας με CycleGAN (**Πηγή**: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks)

3.10: Σύγκριση μεταξύ μοντέλου πολλαπλών τομέων και StarGAN (**Πηγή**: StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation)

3.11: Επισκόπηση του StarGAN, που αποτελείται από δύο ενότητες, έναν διευκρινιστή D και έναν γεννήτορα G (**Πηγή**: StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation)

3.12: Αποτελέσματα StarGAN στο CelebA σύνολο δεδομένων με μεταφορά γνώσης από το RaFD (**Πηγή**: StarGAN: Unified Generative Adversarial Networks
for Multi-Domain Image-to-Image Translation)

4.1: Δείγμα εικόνων από την RAF-DB (**Πηγή**:
http://www.whdeng.cn/raf/model1.html#dataset)

4.2: Δείγμα εικόνων από την AffectNet (**Πηγή**: http://mohammadmahoor.com/affectnet/)

4.3: Πειραματικά αποτελέσματα στο σύνολο επικύρωσης (validation set) για την κατηγοριοποίηση σε οκτώ εκφράσεις (**Πηγή**: http://mohammadmahoor.com/affectnet/)

4.4: Πειραματικά αποτελέσματα στο σύνολο επικύρωσης (validation set) για την κατηγοριοποίηση βάσει σθένους-διέγερσης (**Πηγή**:
http://mohammadmahoor.com/affectnet/)

4.5: Συνολικός αριθμός των εικόνων με χειρόγραφα σχόλια στα σύνολα εκπαίδευσης και επικύρωσης σε κάθε κατηγορία συναισθημάτων (**Πηγή**:
http://mohammadmahoor.com/affectnet/)

4.6: Καρέ της Aff-Wild2 (**Πηγή**: https://ibug.doc.ic.ac.uk/resources/aff-wild2/)

5.1: Γραφική αναπαράσταση υπερεκπαίδευσης και υποεκπαίδευσης (**Πηγή**: paperspace)

5.2: Επίδραση διαφόρων ρυθμών μάθησης στη σύγκλιση (convergence) (**Πηγή**: towardsdatascience)

5.3: Παράδειγμα επαύξησης δεδομένων (**Πηγή**:
https://albumentations.ai/docs/introduction/image_augmentation/)

5.4: Πρώτα αποτελέσματα παραγόμενων εικόνων από το στάδιο δοκιμής.

6.1: Παράδειγμα εξόδων του διευκρινιστή για μία εικόνα.

6.2: Εικόνες που ο διευκρινιστής αξιολόγησε εσφαλμένα ως ψεύτικες.

6.3: Εικόνες που ο διευκρινιστής πραγματοποίησε λάθος κατηγοριοποίηση συναισθήματος.

6.4: Εικόνες που ο διευκρινιστής αξιολόγησε επιτυχώς ως ψεύτικες.

6.5: Εικόνες γεννήτορα που ο διευκρινιστής αξιολόγησε ως αληθινές και κατηγοριοποίησε σωστά το συναίσθημα.

6.6: Εικόνες γεννήτορα για τις οποίες ο διευκρινιστής πραγματοποίησε λάθος κατηγοριοποίηση συναισθήματος.

6.7: Παραδείγματα εικόνων που έχουν δημιουργηθεί από τον γεννήτορα.

6.8: Εικόνες στις οποίες το μικρότερο μέσο τετραγωνικό σφάλμα αντιστοιχούσε στη σωστή αρχική εικόνα.

6.9: Εικόνες στις οποίες το μικρότερο μέσο τετραγωνικό σφάλμα δεν αντιστοιχούσε στη σωστή αρχική εικόνα.

# Κεφάλαιο 1. Εισαγωγή

## 1.1 Αντικείμενο Εργασίας

Σκοπός της παρούσας Διπλωματικής Εργασίας είναι η προσέγγιση της διαδικασίας της Μετάφρασης-Εικόνας-σε-Εικόνα όσον αφορά τον τομέα των ανθρώπινων συναισθημάτων. Η προσέγγιση αυτή περιλαμβάνει τη χρήση Ανταγωνιστικών Γεννητικών Δικτύων τα οποία βασίζονται σε τεχνικές Βαθιάς Μάθησης. Μέρος της εργασίας αποτελεί η μελέτη των Γεννητικών Ανταγωνιστικών Δικτύων και της διαδικασίας της Μετάφρασης Εικόνων, αποσκοπώντας στη βαθύτερη κατανόησή τους ώστε τελικά να γίνει η εκπαίδευση και αξιολόγηση του τελικού μοντέλου και των αποτελεσμάτων.

## 1.2 Αναδρομή

Ανέκαθεν η ανάλυση των ανθρώπινων συναισθημάτων αποτελούσε έναν πολύπλοκο τομέα, με την επιστήμη της Ψυχολογίας να αναπτύσσεται συνεχώς και να κάνει μεγάλη πρόοδο στην ανάλυση και κατανόησή τους. Συνεπώς, ειδικά πριν από μερικά χρόνια, φάνταζε αδύνατο το ενδεχόμενο εκτέλεσης αυτής της διαδικασίας, και πόσο μάλλον η διαδικασία της σύνθεσης των συναισθημάτων, από έναν υπολογιστή.

Η τεχνολογία αναγνώρισης συναισθημάτων (emotion recognition technology) είναι ένας τύπος τεχνητής νοημοσύνης που σχετίζεται με την αναγνώριση του προσώπου και προσπαθεί να προσδιορίσει πώς αισθάνεται ένα άτομο με βάση τις εκφράσεις του προσώπου του και τις σωματικές του ενδείξεις. Στην περίπτωση οπτικοακουστικού υλικού ή κινούμενης εικόνας (π.χ. βίντεο) το λογισμικό μπορεί επίσης να παρακολουθεί επιπλέον παράγοντες όπως την κίνηση των ματιών για να εντοπίζει τα μέρη των ερεθισμάτων στα οποία το άτομο δίνει τη μεγαλύτερη προσοχή [1].

Το ενδιαφέρον για τεχνητή αναγνώριση των συναισθημάτων του προσώπου αυξάνεται όλο και περισσότερο καθώς αναπτύσσονται νέοι αλγόριθμοι και μέθοδοι. Η πρόσφατη πρόοδος στη μηχανική μάθηση έφερε επαναστατικές ανακαλύψεις στον τομέα της έρευνας, και όλο και πιο ακριβή συστήματα αναδύονται κάθε χρόνο. Ωστόσο, αν και η πρόοδος είναι σημαντική, η ανάλυση και κατανόηση των συναισθημάτων εξακολουθεί να είναι μια πολύ μεγάλη πρόκληση [2].

Τα τελευταία χρόνια και κυρίως από το 2014 και μετά, αρκετές μέθοδοι έχουν εφαρμοστεί για την αντιμετώπιση αυτού του δύσκολου αλλά σημαντικού προβλήματος. Πριν το 2014, οι πρώτες παραδοσιακές μέθοδοι στόχευαν στον χειροκίνητο σχεδιασμό χαρακτηριστικών, εμπνευσμένων από ψυχολογικές και νευρολογικές θεωρίες. Τα χαρακτηριστικά περιλάμβαναν το χρώμα, την υφή, τη σύνθεση, την έμφαση, την ισορροπία και άλλα [3, 4]. Επίσης σε αυτό το χρονικό διάστημα πραγματοποιήθηκαν οι πρώτες απόπειρες σύνθεσης ανθρώπινων εκφράσεων με χρήση υπολογιστικών συστημάτων [5, 6].

Οι πρώτες προσπάθειες που επικεντρώθηκαν σε ένα περιορισμένο σύνολο συγκεκριμένων χαρακτηριστικών απέτυχαν να καλύψουν όλους τους σημαντικούς συναισθηματικούς παράγοντες και δεν πέτυχαν επαρκή αποτελέσματα σε μεγάλης κλίμακας σύνολα δεδομένων. Όπως είναι αναμενόμενο, οι σύγχρονες μέθοδοι βαθιάς μάθησης ξεπερνούν τις παραδοσιακές μεθόδους υπολογιστικής όρασης.

Μετά το 2014, ο τομέας της μηχανικής εκμάθησης παρουσιάζει μεγάλη πρόοδο και εισάγονται οι αλγόριθμοι βαθιάς μάθησης [7, 8, 9]. Αυτοί βασίζονται σε μοντέλα νευρωνικών δικτύων όπου συνδεδεμένα επίπεδα νευρώνων χρησιμοποιούνται για την επεξεργασία δεδομένων παρόμοια με τον ανθρώπινο εγκέφαλο. Τα πολλαπλά κρυφά επίπεδα είναι η βάση των νευρωνικών δικτύων για την ανάλυση των λειτουργιών των δεδομένων στο πλαίσιο της λειτουργικής ιεραρχίας. Τα συνελικτικά νευρωνικά δίκτυα (convolutional neural networks — CNN) είναι η πιο δημοφιλής μορφή νευρωνικών δικτύων για την επεξεργασία εικόνων. Το CNN γενικά επιτυγχάνει καλά αποτελέσματα σε εργασίες μηχανικής μάθησης για αναγνώριση συναισθημάτων.

Κυρίως από το 2020 οι πιο σύγχρονες έρευνες στοχεύουν περισσότερο στην παραγωγή ανθρώπινων συναισθημάτων, τα οποία συμβάλλουν και στο πρόβλημα της αναγνώρισης συναισθήματος [10, 11].

Οι έρευνες αυτές και γενικά οι τεχνητή νοημοσύνη για ανάλυση συναισθημάτων έχει ποικίλες εφαρμογές. Για παράδειγμα στον τομέα της υγείας, οι πάροχοι υγειονομικής περίθαλψης μπορούν να χρησιμοποιούν αυτές τις τεχνολογίες για να δίνουν προτεραιότητα στους ασθενείς τους αναλύοντας τις εκφράσεις του προσώπου στην αίθουσα αναμονής, ιδιαίτερα σε επείγοντα κέντρα περίθαλψης όπου οι άνθρωποι δεν προγραμματίζουν ραντεβού. Εκείνοι που αισθάνονται μεγαλύτερη δυσφορία θα μπορούν να λάβουν την υψηλότερη προτεραιότητα, ενώ οι υπόλοιποι είναι σε θέση να περιμένουν για κάποια ώρα.

Επιπλέον, οι ερευνητές χρησιμοποιούν τις τεχνολογίες αναγνώρισης συναισθημάτων και με πιο πειραματικούς τρόπους. Συγκεκριμένα, χρησιμοποιώντας το Google Glass και μια προσαρμοσμένη εφαρμογή, οι ερευνητές της Ιατρικής Σχολής του Πανεπιστημίου του Στάνφορντ βρήκαν έναν τρόπο για να βοηθήσουν τα αυτιστικά παιδιά να αναγνωρίζουν καλύτερα τα συναισθήματα και τις εκφράσεις του προσώπου που συναντούσαν. Η εφαρμογή δίνει στο παιδί ανατροφοδότηση σε πραγματικό χρόνο για τις εκφράσεις των άλλων ανθρώπων, εφόσον το παιδί φοράει το Google Glass [12, 13].

## 1.3 Δομή Εργασίας

Η παρούσα διπλωματική αποσκοπεί στην ανάλυση και σύνθεση συναισθημάτων από εικόνες. Συνεπώς αρχικά στο Κεφάλαιο 2 παρουσιάζουμε τις δομές για την διάκριση των συναισθημάτων ώστε να επιλέξουμε αυτή στην οποία θα βασιστούμε.

Στη συνέχεια, στο Κεφάλαιο 3, παρουσιάζεται το θεωρητικό υπόβαθρο της Βαθιάς Μάθησης και στη συνέχεια, πιο συγκεκριμένα των Γεννητικών Ανταγωνιστικών Δικτύων. Έπειτα αναλύουμε την τεχνική της Μετάφρασης-Εικόνας-Σε-Εικόνα και τα μοντέλα που έχουν αναπτυχθεί, ώστε να καταλήξουμε στο μοντέλο που θα χρησιμοποιηθεί.

Στο  Κεφάλαιο 4 παρουσιάζουμε τα διαθέσιμα σύνολα δεδομένων που υπάρχουν διαθέσιμα και αφορούν τα ανθρώπινα συναισθήματα.

Η πειραματική διαδικασία, οι παράμετροι και η διαδικασία της εκπαίδευσης αναλύεται στο Κεφάλαιο 5.

Τέλος, στο Κεφάλαιο 6 παρουσιάζονται τα αποτελέσματα και γίνεται αξιολόγησή τους.

# Κεφάλαιο 2. Δομές Συναισθημάτων

## 2.1 Εισαγωγή

Η παρούσα διατριβή, όπως αναφέραμε, πραγματεύεται την ανάλυση συναισθημάτων σε εικόνες. Για να επιτευχθεί αυτό χρειάζεται να χρησιμοποιήσουμε κάποια κατηγοριοποίηση των συναισθημάτων ώστε να μπορούμε να βάλουμε ετικέτες (labels) στις εικόνες που χρησιμοποιούμε και αυτές που συνθέτουμε.

## 2.2 Αναδρομή

Η ανάγκη αυτή για κατηγοριοποίηση ωστόσο προήλθε αρκετά νωρίτερα. Ήδη από τον 4ο αιώνα π.Χ., ο Αριστοτέλης προσπάθησε να προσδιορίσει τον ακριβή αριθμό των βασικών συναισθημάτων στον άνθρωπο. Περιγραφόμενος ως ο Κατάλογος των συναισθημάτων του Αριστοτέλη, ο φιλόσοφος πρότεινε 14 διακριτές συναισθηματικές εκφράσεις: φόβος, αυτοπεποίθηση, θυμός, φιλία, ηρεμία, εχθρότητα, ντροπή, αδιαντροπιά, οίκτο, καλοσύνη, φθόνο, αγανάκτηση, μίμηση και περιφρόνηση.

Στη δημοσίευσή του το 1872 “Η Έκφραση των Συναισθημάτων στον Άνθρωπο και τα Ζώα”, ο Κάρολος Δαρβίνος διατύπωσε τη θεωρία ότι τα συναισθήματα ήταν έμφυτα, εξελίχθηκαν και είχαν ένα λειτουργικό σκοπό. Αν και ο Δαρβίνος δεν όρισε ρητά αυτά τα "βασικά συναισθήματα", θεωρείται ότι οραματίστηκε μια μικρότερη λίστα βασικών συναισθημάτων, συμπεριλαμβανομένου του φόβου, του θυμού, της θλίψης, της ευτυχίας και της αγάπης.

Μέχρι τον 20ό αιώνα, με την έλευση της ψυχοθεραπείας, ο αριθμός είχε αυξηθεί σημαντικά. Σύμφωνα με τον Robert Plutchick [14], επίτιμο καθηγητή στο Κολέγιο Ιατρικής του Άλμπερτ Αϊνστάιν, πάνω από 90 διαφορετικοί ορισμοί του "συναισθήματος" έχουν διατυπωθεί από ψυχολόγους με στόχο την ακριβή περιγραφή του τι συνιστά και διαφοροποιεί το ανθρώπινο συναίσθημα [15].

Τα τελευταία χρόνια, οι ψυχολόγοι προσπάθησαν να εντοπίσουν και να κατηγοριοποιήσουν αυτά τα συναισθήματα με έναν τρόπο που θεωρείται εμπειρικός και καθολικός. Ωστόσο, ο αριθμός των συναισθημάτων εξαρτάται σε μεγάλο βαθμό από το πόσο συγκεκριμένα τα συναισθήματα ορίζονται και τα κριτήρια που χρησιμοποιούνται.

Παρακάτω θα γίνει παρουσίαση των τριών βασικών δομών κατηγοριοποίησης που επικρατούν σήμερα και που λάβαμε υπόψη.

## 2.3 Επτά Βασικά Συναισθήματα (Seven Basic Expressions)

### 2.3.1 Κατηγοριοποίηση

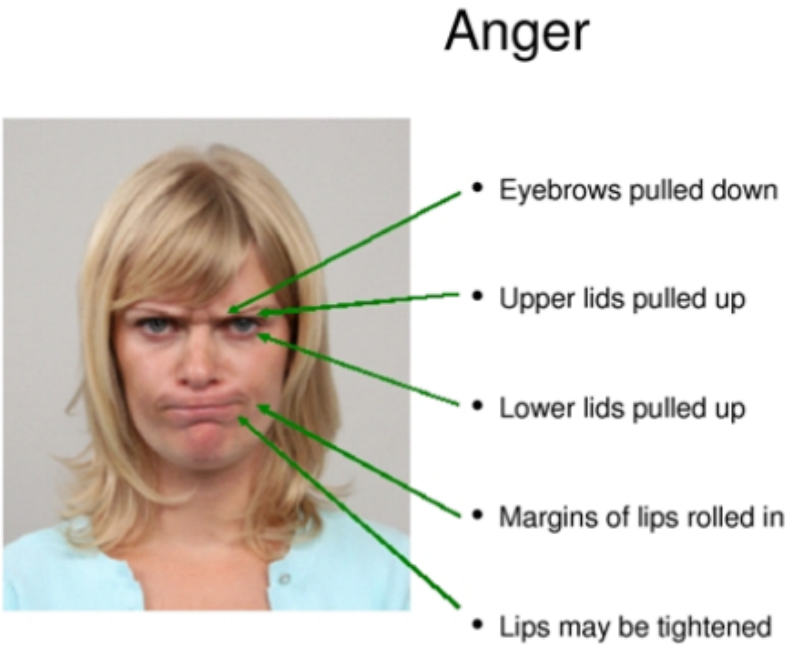
Αργότερα ο ψυχολόγος Paul Ekman προσπάθησε να καθορίσει τα βασικά και καθολικά συναισθήματα που υπάρχουν. Κατά τη διάρκεια της έρευνάς του επάνω στα κύρια συναισθήματα, έδειξε σε ανθρώπους στη Χιλή, στην Αργεντινή, στις ΗΠΑ, στη Βραζιλία και στην Ιαπωνία, εικόνες από πρόσωπα που εξέφραζαν διάφορα συναισθήματα. Σε κάθε περίπτωση οι εξεταζόμενοι είδαν και ταξινόμησαν αυτές τις εικόνες με τον ίδιο ακριβώς τρόπο. Αυτό απέδειξε ότι τα συναισθήματα και η έκφρασή τους δεν είναι κοινωνικά ή πολιτισμικά εξαρτώμενα αλλά είναι περισσότερο αποτέλεσμα εξειδικευμένης βιολογικής σύνδεσης με τον εγκέφαλο.

Με βάση αυτή τη θεωρία του, ο Ekman πρότεινε ότι υπάρχουν επτά συναισθηματικές εκφράσεις καθολικές για τους ανθρώπους σε όλο τον κόσμο: χαρά, θλίψη, έκπληξη, φόβος, θυμός, αηδία και περιφρόνηση [16].

### 2.3.2 Ανάλυση

Στη συνέχεια θα δούμε αναλυτικά τα συναισθήματα που ανήκουν στα επτά βασικά καθώς και τα χαρακτηριστικά του προσώπου που συνθέτουν το καθένα από αυτά.

Αρχικά είναι ο θυμός. Ένα θυμωμένο πρόσωπο χαρακτηρίζεται από τα εξής: κατεβασμένα φρύδια, άνω και κάτω βλέφαρα κατεβασμένα, χείλη γυρισμένα προς τα μέσα και σφιχτά.

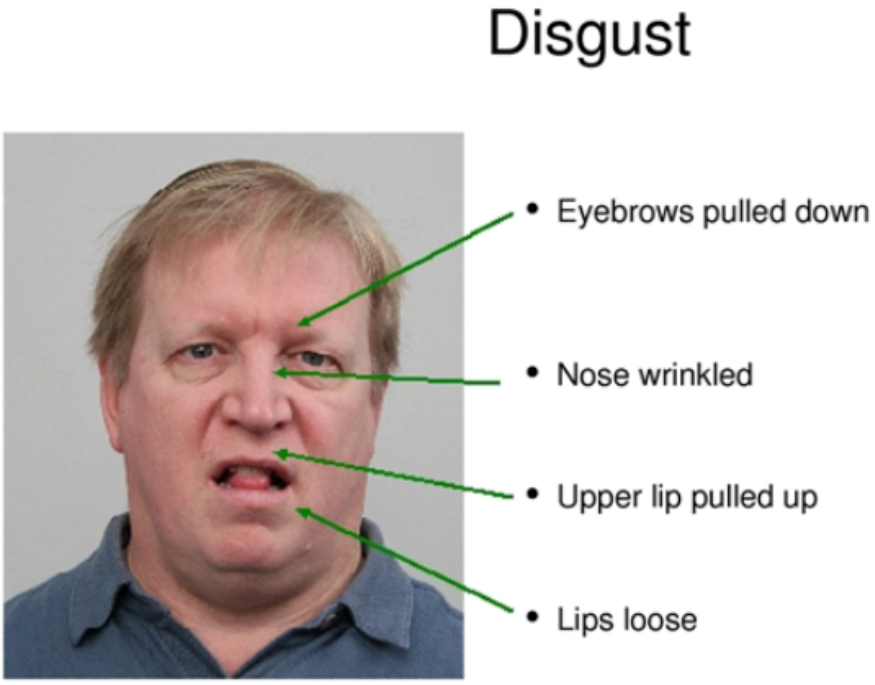



Σχήμα 2.1: Χαρακτηριστικά του θυμού βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

Δεύτερο βασικό συναίσθημα είναι ο φόβος κατά τον οποίο τα φρύδια και τα άνω βλέφαρα είναι σηκωμένα και το στόμα ανοιχτό. Παρεμφερή συναισθήματα όπως π.χ. ο τρόμος περιλαμβάνουν επίσης αυτά τα χαρακτηριστικά.

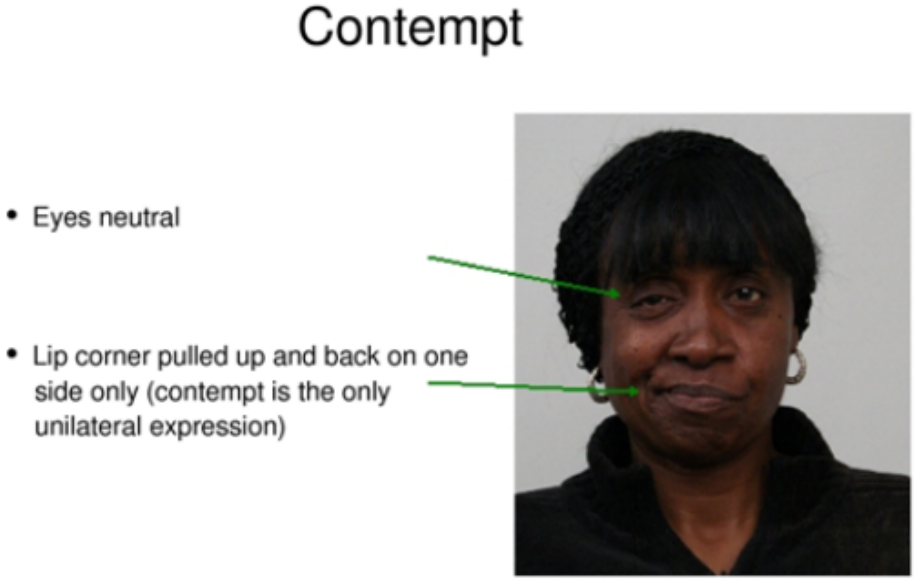

Σχήμα 2.2: Χαρακτηριστικά του φόβου βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

Επόμενο συναίσθημα είναι η αηδία όπου τα φρύδια είναι κατεβασμένα, η μύτη ζαρωμένη και τα χείλη χαλαρά με το άνω να είναι σηκωμένο.



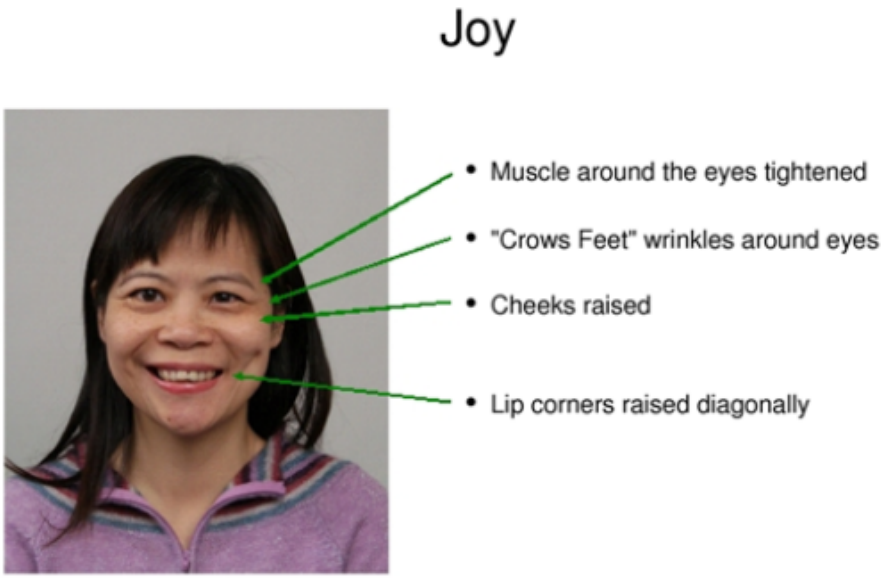

Σχήμα 2.3: Χαρακτηριστικά της αηδίας βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

Κατά την περιφρόνηση τα μάτια είναι ουδέτερα και η άκρη του χείλους σηκωμένη άλλα μόνο από την μία πλευρά. Είναι αξιοσημείωτο ότι η περιφρόνηση είναι το μοναδικό συναίσθημα κατά το οποίο συμβαίνει αυτό, καθώς σε όλες τις υπόλοιπες εκφράσεις τα χαρακτηριστικά είναι συμμετρικά. Ωστόσο πολλές φορές η περιφρόνηση δεν περιλαμβάνεται στα βασικά συναισθήματα. Συγκεκριμένα ο James Russell σε μία έρευνα το 1991 αποφάνθηκε ότι το ανασηκωμένο χείλος είναι ανεπαρκές για την αναγνώριση της αυτού του συναισθήματος.



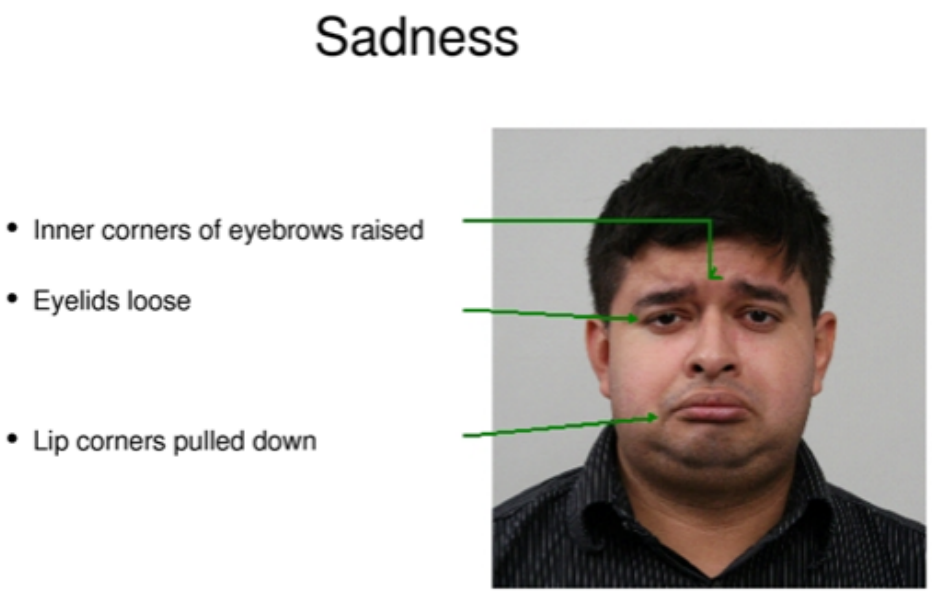

Σχήμα 2.4: Χαρακτηριστικά της περιφρόνησης βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

Τα χαρακτηριστικά πάνω στο πρόσωπο που υποδηλώνουν χαρά είναι ότι οι μύες γύρω από τα μάτια είναι σφιχτοί, δημιουργούνται ρυτίδες γύρω από τα μάτια, τα μάγουλα είναι ανασηκωμένα καθώς και οι άκρες των χειλιών διαγωνίως.



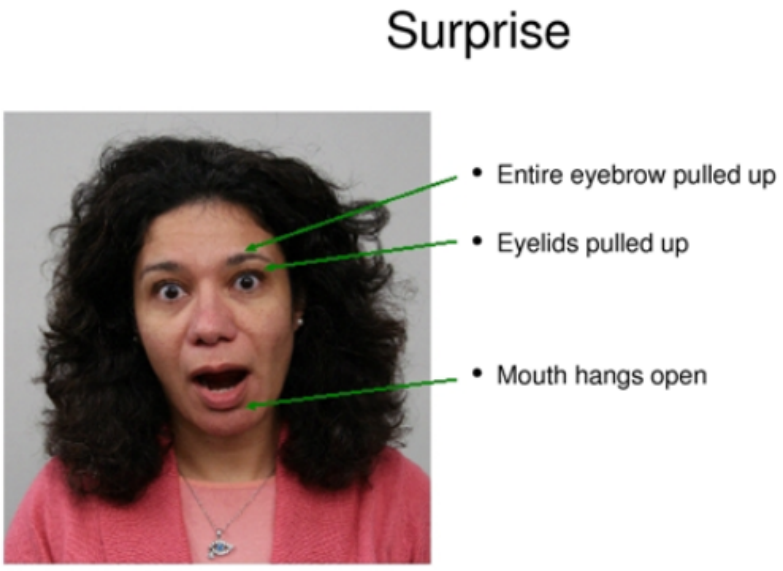

Σχήμα 2.5: Χαρακτηριστικά της χαράς βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

Στη θλίψη οι εσωτερικές γωνίες των φρυδιών είναι ανασηκωμένες, τα βλέφαρα χαλαρά και οι γωνίες των χειλιών κατεβασμένες.




Σχήμα 2.6: Χαρακτηριστικά της θλίψης βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

Τέλος, η έκπληξη αντικατοπτρίζεται στο πρόσωπο με ανασηκωμένα και τα δύο φρύδια και τα βλέφαρα καθώς και το στόμα ανοιχτό.




Σχήμα 2.7: Χαρακτηριστικά της έκπληξης βάσει των επτά βασικών συναισθημάτων (**Πηγή**: Humintell)

Ένα επιπλέον συναίσθημα που περιλαμβάνεται συχνά σε αυτή την κατηγοριοποίηση είναι το ουδέτερο όπου η έκφραση είναι κενή και υπονοεί έλλειψη αισθητού συναισθήματος. Τις

περισσότερες φορές ένα ουδέτερο πρόσωπο ορίζεται από εντελώς ευθείο στόμα, μη εστιασμένα μάτια, και χαλαρά μάγουλα.

Συναισθήματα όπως η ντροπή, η υπερηφάνεια, η ζήλια και η ενοχή αν και είναι σημαντικά, δεν θεωρούνται μέρος των βασικών συναισθημάτων. Συγκεκριμένα δεν υπάρχουν επιστημονικά στοιχεία που να δείχνουν ότι υπάρχει μια καθολική έκφραση ντροπής στο πρόσωπο. Ωστόσο, πρόσφατες έρευνες έχουν δείξει ότι υπάρχουν κινήσεις του προσώπου και του σώματος που μπορεί να είναι καθολικές για θρίαμβο, ντροπή και αμηχανία.

## 2.4 Μονάδες Δράσης Προσώπου (Facial Action Units)

### 2.4.1 Κατηγοριοποίηση

Ο Ekman μαζί με τον Wallace Friesen ανέπτυξαν το ‘Σύστημα Κωδικοποίησης της Δράσης του Προσώπου’ (Facial Action Coding System, FACS) για την περιγραφή των εκφράσεων του προσώπου από μονάδες δράσης (action units, AU). Συγκεκριμένα, οι μονάδες δράσης (Action Units, AU) είναι οι θεμελιώδεις ενέργειες των μεμονωμένων μυών ή των ομάδων μυών. Από τις 44 FACS AU που όρισαν, οι 30 AU σχετίζονται ανατομικά με τις συσπάσεις συγκεκριμένων μυών του προσώπου: 12 είναι για το άνω πρόσωπο, και 18 είναι για το κάτω πρόσωπο. Οι AU μπορούν να εμφανιστούν είτε μεμονωμένα είτε σε συνδυασμό.Αν και ο αριθμός των μονάδων ατομικής δράσης είναι σχετικά μικρός, έχουν παρατηρηθεί περισσότεροι από 7.000 διαφορετικοί συνδυασμοί AU [17, 18, 19].

### 2.4.2 Ανάλυση

Χρησιμοποιώντας το FACS μπορεί να κωδικοποιηθεί σχεδόν οποιαδήποτε ανατομικά πιθανή έκφραση του προσώπου, γίνοντας αποσύνθεσή της σε συγκεκριμένες AU και τα τμήματα τους που παρήγαγαν την έκφραση. Δεδομένου ότι οι AU είναι ανεξάρτητες από οποιαδήποτε ερμηνεία, μπορούν να χρησιμοποιηθούν για οποιαδήποτε διαδικασία λήψης αποφάσεων υψηλότερης τάξης, συμπεριλαμβανομένης της αναγνώρισης βασικών συναισθημάτων.

Το FACS πέρα από τις AU, οι οποίες όπως αναφέρθηκε είναι μια συστολή ή χαλάρωση ενός ή περισσότερων μυών, ορίζει επίσης μια σειρά από περιγραφείς δράσεων (Action Descriptors), οι οποίοι είναι κινήσεις που μπορεί να περιλαμβάνουν τις ενέργειες αρκετών μυϊκών ομάδων (π.χ., μια εμπρόσθια κίνηση της γνάθου).

Αν και η ονομάτιση των εκφράσεων απαιτεί επί του παρόντος εκπαιδευμένους ειδικούς, οι ερευνητές σημείωσαν κάποια επιτυχία στη χρήση υπολογιστών για τον αυτόματο εντοπισμό των κωδικών FACS. Κάποια υπολογιστικά μοντέλα επιτρέπουν στις εκφράσεις να τίθενται τεχνητά θέτοντας τις επιθυμητές μονάδες δράσης.

Ουσιαστικά στον FACS έχει γίνει συγκεκριμένη αρίθμηση των AUs και κάθε νούμερο αντιστοιχεί σε μία συγκεκριμένη δράση (π.χ. AU 0 : ουδέτερο πρόσωπο, AU 1 : ανασήκωση του εσωτερικού μέρους του φρυδιού κ.ο.κ.).

| Action Unit # | Action |
|---|---|
| AU 1 | inner brow raiser |
| AU 2 | outer brow raiser |
| AU 4 | brow lowerer |
| AU 6 | cheek raiser |
| AU 12 | lip corner puller |
| AU 15 | lip corner depressor |
| AU 20 | lip stretcher |
| AU 25 | lips part |

Σχήμα 2.8: Παράδειγμα αντιστοίχισης των AU με τις δράσεις τους (**Πηγή**: https://arxiv.org/pdf/2001.11409.pdf )

Οι εντάσεις του FACS σχολιάζονται με την προσθήκη γραμμάτων A-E (για ελάχιστη - μέγιστη ένταση) στον αριθμό μονάδας δράσης (π.χ. η AU 1A είναι το πιο αδύναμο ίχνος της AU 1 και η AU 1E είναι η μέγιστη δυνατή ένταση για το κάθε άτομο). Συγκεκριμένα:

- A : Ίχνος
- B : Ελαφρώς
- C : Αρκετά
- D : Σοβαρά
- E : Μέγιστο

Υπάρχουν και άλλες σημάνσεις που είναι παρούσες στους κώδικες FACS για τις συναισθηματικές εκφράσεις, όπως το "R" που αντιπροσωπεύει μια δράση που εμφανίζεται στη δεξιά πλευρά του προσώπου και το "L" για τις ενέργειες που συμβαίνουν στα αριστερά. Μια ενέργεια που είναι μονομερής (εμφανίζεται μόνο στη μία πλευρά του προσώπου) αλλά δεν έχει συγκεκριμένη πλευρά υποδεικνύεται με ένα "U" και μια ενέργεια που είναι μονομερής αλλά έχει μια ισχυρότερη πλευρά υποδεικνύεται με ένα "A".

Έτσι βάσει όλων των παραπάνω μπορούμε να συνθέσουμε όλα τα βασικά συναισθήματα (π.χ. Χαρά : AU 6 + AU 12 όπου AU 6 : μάγουλα ανασηκωμένα και AU 12 : άκρη του χείλους ανασηκωμένη) [20].

| Emotion | Action units |
|---|---|
| Happiness | 6+12 |
| Sadness | 1+4+15 |
| Surprise | 1+2+5B+26 |
| Fear | 1+2+4+5+7+20+26 |
| Anger | 4+5+7+23 |
| Disgust | 9+15+17 |
| Contempt | R12A+R14A |

Σχήμα 2.9: Τα επτά βασικά συναισθήματα με σύνθεση μονάδων δράσης (**Πηγή**: Wikipedia)

## 2.5 Σθένος και διέγερση (Valence and arousal)

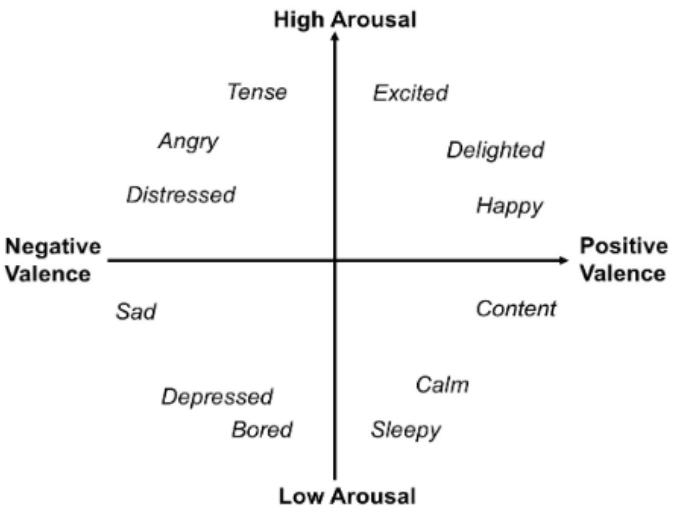
### 2.5.1 Κατηγοριοποίηση

Ο James Russell το 1980 διατύπωσε την αντίληψη ότι το κάθε συναίσθημα έχει δύο διαστάσεις. Το μοντέλο αυτό προτείνει ότι όλες οι συναισθηματικές καταστάσεις προκύπτουν από τις γνωστικές ερμηνείες των νευρικών αισθήσεων του πυρήνα που είναι το προϊόν δύο ανεξάρτητων νευροφυσιολογικών συστημάτων. Αυτό το μοντέλο βρίσκεται σε αντίθεση με τις θεωρίες των βασικών συναισθημάτων, οι οποίες θεωρούν ότι ένα διακριτό και ανεξάρτητο νευρικό σύστημα προάγει κάθε συναίσθημα. [21, 22].

### 2.5.2 Ανάλυση

Σε αυτό το δισδιάστατο μοντέλο μία διάσταση, η οριζόντια, είναι το σθένος, δηλαδή πόσο θετικό ή αρνητικό είναι το συναίσθημα (π.χ. όταν κάποιος χαμογελάει είναι πιο δεξιά στην κλίμακα ενώ όταν κάποιος κλαίει πιο αριστερά). Η δεύτερη διάσταση, η κάθετη, είναι η διέγερση, δηλαδή πόσο συνεπαίρνει ένα συναίσθημα (π.χ. όταν κάποιος είναι νυσταγμένος είναι χαμηλά στη κλίμακα της διέγερσης ενώ όταν είναι ενθουσιασμένος για κάτι υψηλά).

Οι δύο διαστάσεις μπορούν να χαρτογραφηθούν σε ένα δισδιάστατο χώρο και οι συνδυασμοί των διαφορετικών αξιών της σθένους και της διέγερσης συνδέονται με διαφορετικά διακριτά συναισθήματα. Δηλαδή κάθε συναίσθημα μπορεί να κατανοηθεί ως ένας γραμμικός συνδυασμός αυτών των δύο διαστάσεων, ή ως διαφορετικοί βαθμοί σθένους και διέγερσης, με συνεχείς τιμές σε κάθε άξονα στο διάστημα [-1,1]. Παραδείγματος χάριν, η επάνω αριστερή γωνία του δισδιάστατου χώρου αντιπροσωπεύει συναισθήματα αρνητικής σθένους και υψηλής διέγερσης, όπως ο θυμός. η κάτω δεξιά γωνία του χώρου αντιπροσωπεύει συναισθήματα θετικής σθένους και χαμηλής διέγερσης, όπως η ηρεμία.




Σχήμα 2.10: Η δισδιάστατη δομή των συναισθημάτων σύμφωνα με τον Russell (**Πηγή**: https://arxiv.org/ftp/arxiv/papers/2001/2001.04509.pdf)

Η χαρά, για παράδειγμα, εννοείται ως μια συναισθηματική κατάσταση που είναι το προϊόν ισχυρής ενεργοποίησης στα νευρικά συστήματα που σχετίζονται με θετικό σθένος ή ευχαρίστηση μαζί με μέτρια ενεργοποίηση στα νευρικά συστήματα που σχετίζονται με διέγερση. Συναισθηματικές καταστάσεις εκτός από τη χαρά προκύπτουν επίσης από τα ίδια δύο νευροφυσιολογικά συστήματα, αλλά διαφέρουν στο βαθμό ή την έκταση της ενεργοποίησης. Συγκεκριμένα συναισθήματα προκύπτουν επομένως από μοτίβα ενεργοποίησης μέσα σε αυτά τα δύο νευροφυσιολογικά συστήματα, μαζί με γνωστικές ερμηνείες και επισήμανση αυτών των βασικών φυσιολογικών εμπειριών.

## 2.6 Συμπεράσματα

Λαμβάνοντας υπ'όψιν τις παραπάνω δομές καταλήξαμε στην επιλογή της πρώτης, δηλαδή στα 'Επτά βασικά Συναισθήματα' για τους σκοπούς αυτής της διατριβής. Αυτή η επιλογή έγινε καθώς η συγκεκριμένη δομή είναι η πιο διαδεδομένη, χρησιμοποιείται κατα κόρον και είναι πιο καλά ορισμένη από τις υπόλοιπες οπότε υπάρχουν και διαθέσιμες προς χρήση βάσεις δεδομένων. Στην κατηγοριοποίηση μας δεν συμπεριλαμβάνουμε την καταφρόνηση αλλά την ουδέτερη έκφραση για λόγους ευκολίας.

Αντιθέτως, όσον αφορά το μοντέλο των μονάδων δράσης του προσώπου δεν υπάρχουν πολλές διαθέσιμες βάσεις καθώς η διάκριση των εκφράσεων απαιτεί ειδικούς οι οποίοι έχουν λάβει ειδική εκπαίδευση και έχουν πιστοποίηση.

Δεν επικεντρωθήκαμε στο μοντέλο του σθένους-διέγερσης επειδή εμπεριέχει έναν βαθμό υποκειμενικότητας, καθώς οι τιμές για τα συναισθήματα, όπως αναφέρθηκε, παίρνουν συνεχείς τιμές σε κάθε άξονα στο διάστημα [-1,1]. Συνεπώς είναι δύσκολο να καθορίσουμε την ακριβή τιμή (π.χ. αν ένα συναίσθημα στον άξονα της διέγερσης πρέπει να είναι 0.6 ή 0.62)

# Κεφάλαιο 3. Μετάφραση Εικόνας-σε-Εικόνα με χρήση Ανταγωνιστικών Γεννητικών Δικτύων

## 3.1 Εισαγωγή

Στην παρούσα διπλωματική προσπαθούμε να μετατρέψουμε ένα συναίσθημα σε ένα άλλο, διατηρώντας όμως την βασική αναπαράσταση της αρχικής εικόνας, κάνουμε δηλαδή μετάφραση εικόνας-σε-εικόνα. Αυτό το επιτυγχάνουμε με την χρήση Βαθιών Νευρωνικών Δικτύων (Deep Neural Networks) και συγκεκριμενα Ανταγωνιστικών Γεννητικών Δικτύων (Generative Adversarial Networks).

## 3.2 Βαθιά Νευρωνικά Δίκτυα

### 3.2.1 Βαθιά Μάθηση (Deep Learning)

Τα γεννητικά ανταγωνιστικά δίκτυα που θα αναλύσουμε στη συνέχεια χρησιμοποιούν τεχνικές βαθιάς μάθησης.

Τα περισσότερα σύγχρονα μοντέλα βαθιάς μάθησης βασίζονται σε τεχνητά νευρωνικά δίκτυα, ειδικά συνελικτικά νευρωνικά δίκτυα, αν και μπορούν επίσης να περιλαμβάνουν προτασιακούς τύπους ή λανθάνουσες μεταβλητές οργανωμένες με επίπεδα σε βαθιά παραγωγικά μοντέλα, όπως οι κόμβοι σε βαθιά δίκτυα πεποιθήσεων και οι μηχανές Boltzmann.

Στη βαθιά μάθηση, κάθε επίπεδο μαθαίνει να μετατρέπει τα δεδομένα εισόδου του σε μια ελαφρώς πιο αφηρημένη και σύνθετη αναπαράσταση. Σε μια εφαρμογή αναγνώρισης εικόνων, η πρωτογενής είσοδος μπορεί να είναι ένας πίνακας, που παρουσιάζεται με μορφή pixels: η πρώτη στρώση αναπαράστασης μπορεί να αφαιρεί τα πίξελ και να κωδικοποιεί τις ακμές, η δεύτερη στρώση μπορεί να συνθέτει και να κωδικοποιεί διατάξεις ακμών, η τρίτη στρώση μπορεί να κωδικοποιεί την μύτη και τα μάτια και η τέταρτη στρώση μπορεί να αναγνωρίσει ότι η εικόνα περιέχει μια όψη. Είναι σημαντικό ότι μια βαθιά διαδικασία μάθησης μπορεί να μάθει ποιες λειτουργίες να τοποθετήσει καλύτερα σε ποιο επίπεδο από μόνη της. Αυτό ωστόσο δεν εξαλείφει εντελώς την ανάγκη για χειροκίνητη ρύθμιση.

Η λέξη «βαθιά» στην «βαθιά μάθηση» αναφέρεται στον αριθμό των επιπέδων μέσω των οποίων τα δεδομένα μετασχηματίζονται από την είσοδο μέχρι την έξοδο του δικτύου [23, 24].

### 3.2.2 Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks)

Ένα βαθύ νευρωνικό δίκτυο (Deep Neural Network, DNN) είναι ένα τεχνητό νευρωνικό δίκτυο με πολλαπλά στρώματα μεταξύ των επιπέδων εισόδου και εξόδου.Υπάρχουν

διαφορετικοί τύποι νευρωνικών δικτύων αλλά πάντα αποτελούνται από τα ίδια συστατικά: νευρώνες, συνάψεις, βάρη, σταθεροί όροι και συναρτήσεις. Αυτά τα συστατικά λειτουργούν παρόμοια με τον ανθρώπινο εγκέφαλο και μπορούν να εκπαιδευτούν όπως οποιοδήποτε άλλο αλγόριθμο μηχανικής μάθησης (Machine Learning, ML)
Για παράδειγμα, ένα DNN που έχει εκπαιδευτεί να αναγνωρίζει φυλές σκύλων θα πάει πάνω από τη δεδομένη εικόνα και να υπολογίσει την πιθανότητα ότι ο σκύλος στην εικόνα είναι μια συγκεκριμένη φυλή. Ο χρήστης μπορεί να εξετάσει τα αποτελέσματα και να επιλέξει ποιες πιθανότητες θα πρέπει να εμφανίζει το δίκτυο (πάνω από ένα συγκεκριμένο όριο κ.λπ.) και να επιστρέψει την προτεινόμενη ετικέτα. Κάθε τέτοια μαθηματική χειραγώγηση θεωρείται ένα επίπεδο, και τα σύνθετα  DNN έχουν πολλά επίπεδα, εξ ου και το όνομα "βαθιά" δίκτυα.

Τα DNN μπορούν να μοντελοποιούν πολύπλοκες μη γραμμικές σχέσεις. Οι αρχιτεκτονικές DNN παράγουν σύνθετα μοντέλα όπου το αντικείμενο εκφράζεται ως σύνθεση επιπέδων των αρχέτυπων. Τα επιπλέον επίπεδα επιτρέπουν τη σύνθεση χαρακτηριστικών από χαμηλότερα επίπεδα, ενδεχομένως μοντελοποιώντας σύνθετα δεδομένα με λιγότερες μονάδες από ένα παρόμοιο ρηχό δίκτυο.

Οι βαθιές αρχιτεκτονικές περιλαμβάνουν πολλές παραλλαγές μερικών βασικών προσεγγίσεων. Κάθε αρχιτεκτονική έχει βρει επιτυχία σε συγκεκριμένους τομείς. Δεν είναι πάντα εφικτό να συγκρίνουμε την απόδοση πολλών αρχιτεκτονικών, εκτός αν έχουν αξιολογηθεί στα ίδια σύνολα δεδομένων.

Τα DNN συνήθως είναι δίκτυα εμπρόσθιας προώθησης στα οποία τα δεδομένα ρέουν από το επίπεδο εισόδου στο επίπεδο εξόδου χωρίς να χρειάζεται επανάληψη. Αρχικά, το DNN δημιουργεί έναν χάρτη εικονικών νευρώνων και εκχωρεί τυχαίες αριθμητικές τιμές, ή "βάρη", στις συνδέσεις μεταξύ τους. Τα βάρη και οι είσοδοι πολλαπλασιάζονται και επιστρέφουν μια έξοδο μεταξύ 0 και 1. Αν το δίκτυο δεν αναγνωρίζει με ακρίβεια ένα συγκεκριμένο μοτίβο, ο αλγόριθμος μάθησης θα προσαρμόσει τα βάρη. Με αυτόν τον τρόπο ο αλγόριθμος μπορεί να κάνει ορισμένες παραμέτρους πιο σημαντικές, μέχρι να καθορίσει το σωστό μαθηματικό χειρισμό για την πλήρη επεξεργασία των δεδομένων [25, 26].

## 3.3 Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks)

### 3.3.1 Εισαγωγή

Τα Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks, GANs), είναι μια προσέγγιση για παραγωγική μοντελοποίηση με τη χρήση μεθόδων βαθιάς μάθησης, όπως τα συνελικτικα νευρωνικά δίκτυα.

Η γεννητική μοντελοποίηση (generative modeling) είναι μια εργασία μάθησης χωρίς επίβλεψη στη μηχανική μάθηση, η οποία περιλαμβάνει την αυτόματη ανακάλυψη και εκμάθηση μοτίβων στα δεδομένα  εισόδου, με τέτοιο τρόπο ώστε το μοντέλο να μπορεί να χρησιμοποιηθεί για την παραγωγή ή εξαγωγή νέων παραδειγμάτων που πιθανόν να έχουν σχεδιαστεί από το αρχικό σύνολο δεδομένων.

Τα GAN είναι ένας έξυπνος τρόπος εκπαίδευσης ενός παραγωγικού μοντέλου, διαμορφώνοντας το πρόβλημα ως επιβλεπόμενο μαθησιακό πρόβλημα με δύο υπομοντέλα: το μοντέλο του γεννήτορα (generator) που εκπαιδεύουμε για να παράγουμε νέα παραδείγματα, και το μοντέλο του διευκρινιστή (discriminator) που προσπαθεί να ταξινομήσει παραδείγματα είτε ως πραγματικά είτε ως ψευδή (αυτά που παράγονται). Τα δύο μοντέλα εκπαιδεύονται μαζί σε ένα παιχνίδι μηδενικού αθροίσματος, ανταγωνιστικά, μέχρι ο discriminator να ξεγελαστεί περίπου τις μισές φορές, που σημαίνει ότι ο generator παράγει αληθοφανή παραδείγματα.

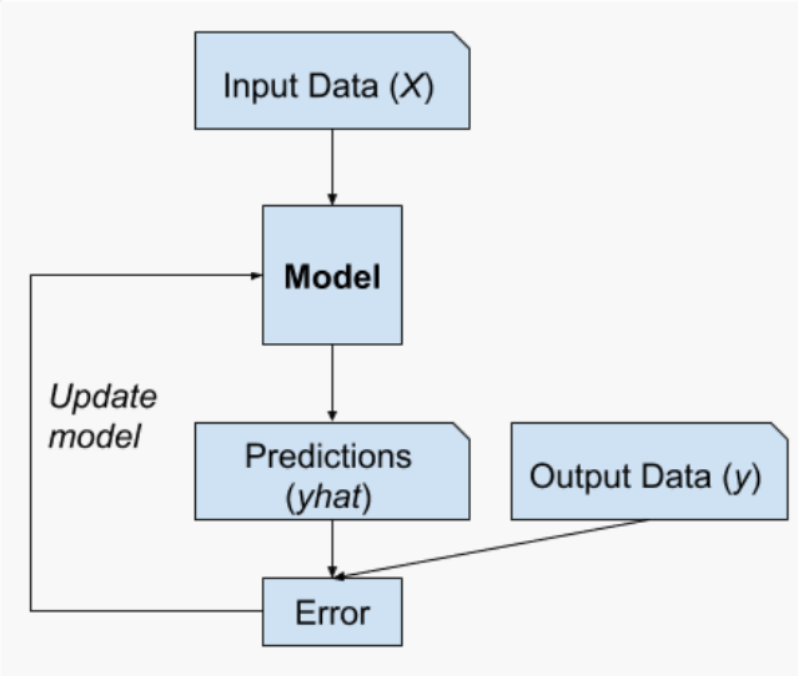
Τα GANs πλέον ανήκουν σε ένα ταχέως μεταβαλλόμενο πεδίο, δίνοντας την δυνατότητα στα γεννητικά μοντέλα να παραγάγουν ρεαλιστικά παραδείγματα σε μια σειρά προβλημάτων, κυρίως σε εργασίες μετάφρασης εικόνων σε εικόνες, όπως π.χ. η μετατροπή φωτογραφιών από καλοκαίρι σε χειμώνα, ή μέρα στη νύχτα, και στην παραγωγή φωτορεαλιστικών φωτογραφιών αντικειμένων, ανθρωπων και σκηνών που ακόμη και οι άνθρωποι δεν μπορούν να αναγνωρίσουν αν είναι ψεύτικες [27, 28].

### 3.3.2 Επιβλεπόμενη και μη-επιβλεπόμενη εκμάθηση

Ένα τυπικό πρόβλημα μηχανικής μάθησης περιλαμβάνει τη χρήση ενός μοντέλου για την πραγματοποίηση μιας πρόβλεψης, δηλ. για προγνωστική μοντελοποίηση.

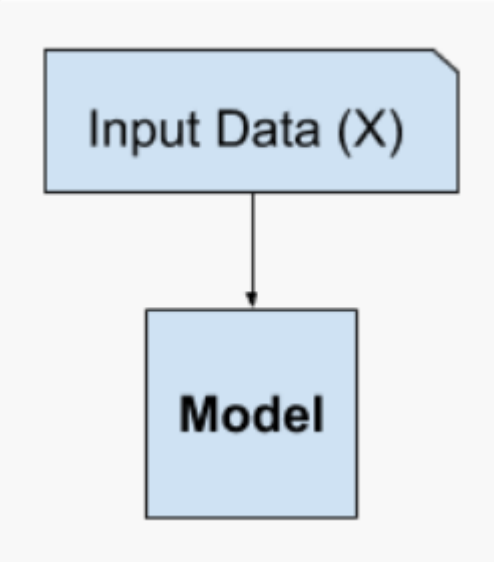
Αυτό απαιτεί ένα σύνολο δεδομένων εκπαίδευσης που χρησιμοποιείται για την εκπαίδευση ενός μοντέλου, το οποίο αποτελείται από πολλά παραδείγματα, τα οποία ονομάζονται δείγματα (samples), καθένα με μεταβλητές εισόδου (X) και ετικέτες κλάσεων εξόδου (y). Ένα μοντέλο εκπαιδεύεται παρουσιάζοντάς του παραδείγματα εισόδου και αφήνοντάς το να προβλέψει τις εξόδους και στη συνέχεια διορθώνοντας το μοντέλο ώστε οι έξοδοι να μοιάζουν περισσότερο με τις αναμενόμενες. Αυτή η διόρθωση του μοντέλου αναφέρεται γενικά ως επιβλεπόμενη μάθηση.

Ουσιαστικά στην προγνωστική (predictive) ή επιβλεπόμενη (supervised learning) προσέγγιση μάθησης, ο στόχος είναι να μάθουμε μια αντιστοίχιση από εισόδους X σε εξόδους y, με δεδομένο ένα σύνολο ζευγών εισόδου-εξόδου.




Σχήμα 3.1: Παράδειγμα επιβλεπόμενης μάθησης (**Πηγή**: machinelearningmastery)
Παραδείγματα επιβλεπόμενων μαθησιακών προβλημάτων είναι η ταξινόμηση και η παλινδρόμηση, και παραδείγματα εποπτευόμενων αλγορίθμων περιλαμβάνουν τη λογιστική παλινδρόμηση και το τυχαίο δάσος.

Ένα άλλο παράδειγμα μάθησης είναι αυτό όπου το μοντέλο λαμβάνει μόνο τις μεταβλητές εισόδου (X) και το πρόβλημα δεν έχει μεταβλητές εξόδου (y). Το μοντέλο κατασκευάζεται με εξαγωγή, ή σύνοψη των μοτίβων στα δεδομένα εισόδου. Δεν υπάρχει διόρθωση του μοντέλου, καθώς το μοντέλο δεν προβλέπει μια συγκεκριμένη έξοδο. Αυτή η έλλειψη διόρθωσης αναφέρεται γενικά ως μη επιβλεπόμενη μορφή μάθησης.




Σχήμα 3.2: Παράδειγμα μη επιβλεπόμενης μάθησης (**Πηγή**: machinelearningmastery)

Παραδείγματα μη-επιβλεπόμενων μαθησιακών προβλημάτων περιλαμβάνουν τη συσταδοποίηση (clustering) και τη γενετική μοντελοποίηση, και παραδείγματα μη-επιβλεπόμενων αλγορίθμων είναι τα K-μέσων (K-means) και τα Γεννητικά Ανταγωνιστικά Δίκτυα [29, 30, 31]

### 3.3.3 Διευκρινιστής και γεννήτορας

Στην επιβλεπόμενη μάθηση μας ενδιαφέρει η ανάπτυξη ενός μοντέλου για την πρόβλεψη της ετικέτας μιας κλάσης δίνοντας ως παράδειγμα κάποιες μεταβλητές εισόδου. Αυτή η εργασία προγνωστικής μοντελοποίησης ονομάζεται ταξινόμηση (classification). Η ταξινόμηση επίσης παραδοσιακά αναφέρεται ως διευκρινιστική μοντελοποίηση.
Αυτό συμβαίνει επειδή το μοντέλο πρέπει να διακρίνει παραδείγματα μεταβλητών εισόδου από κλάσεις σε κλάσεις και πρέπει να επιλέξει, ή να αποφασίσει, ποια κλάση ανήκει σε ένα συγκεκριμένο παράδειγμα.

Εναλλακτικά, μη επιβλεπόμενα μοντέλα που συνοψίζουν την κατανομή των μεταβλητών εισόδου μπορούν να χρησιμοποιηθούν για να δημιουργήσουν, ή να παραγάγουν νέα παραδείγματα στην κατανομή της εισόδου. Ως εκ τούτου, αυτοί οι τύποι των μοντέλων αναφέρονται ως γεννητικά μοντέλα.

Ένα γεννητικό μοντέλο μπορεί να είναι σε θέση να συνοψίσει επαρκώς την κατανομή των δεδομένων, και στη συνέχεια να χρησιμοποιηθεί για να δημιουργήσει νέες μεταβλητές που ταιριάζουν σωστά στην κατανομή της μεταβλητής εισόδου. Μάλιστα, ένα πραγματικά καλό γεννητικό μοντέλο μπορεί να είναι σε θέση να παράγει νέα παραδείγματα που δεν είναι απλώς πιθανά, αλλά επιπλέον δεν ξεχωρίζουν από τα πραγματικά παραδείγματα του προβληματικού τομέα.

Τα GANs είναι ένα γεννητικό μοντέλο βασισμένο σε τεχνικές βαθιάς μάθησης. Γενικότερα, τα GANs είναι μια αρχιτεκτονική για την εκπαίδευση ενός γεννητικού μοντέλου, και είναι πιο συνηθισμένο να χρησιμοποιούνται μοντέλα βαθιάς μάθησης σε αυτήν την αρχιτεκτονική. Η πρώτη αναφορά στα GANs έγινε σε δημοσίευση του 2014 από τον Ian Goodfellow.

Η αρχιτεκτονική του μοντέλου του GAN περιλαμβάνει δύο υπομοντέλα: ένα μοντέλο γεννήτορα για τη δημιουργία νέων παραδειγμάτων και ένα μοντέλο διευκρινιστή για να αποφανθεί αν τα παραγόμενα παραδείγματα είναι πραγματικά, από το σύνολο των δεδομένων, ή ψεύτικα, που παράγονται από το μοντέλο του γεννήτορα.

Ο γεννήτορας δέχεται ως είσοδο ένα τυχαίο διάνυσμα σταθερού μήκους και παράγει ένα δείγμα στο πεδίο ορισμού. Το διάνυσμα παράγεται τυχαία από μια κατανομή Gauss, και χρησιμοποιείται ως seed στην γεννητική διαδικασία. Μετά την εκπαίδευση, τα σημεία σε αυτόν τον πολυδιάστατο διανυσματικό χώρο θα αντιστοιχούν σε σημεία στο πεδίο του προβλήματος, σχηματίζοντας μια συμπιεσμένη αναπαράσταση της κατανομής των δεδομένων.

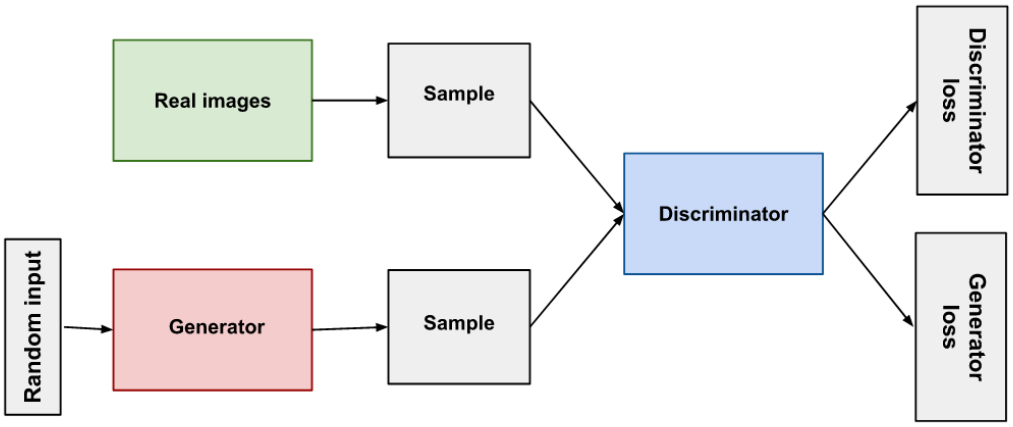
Αυτός ο διανυσματικός χώρος αναφέρεται ως λανθάνων χώρος ή διανυσματικός χώρος που αποτελείται από λανθάνουσες μεταβλητές. Λανθάνουσες μεταβλητές (latent variables), ή κρυφές μεταβλητές, είναι εκείνες οι μεταβλητές που είναι σημαντικές για ένα πεδίο ορισμού, αλλά δεν είναι άμεσα παρατηρήσιμες.

Συχνά αναφερόμαστε σε λανθάνουσες μεταβλητές, ή λανθάνοντα χώρο, ως προβολή ή συμπίεση μιας κατανομής δεδομένων. Δηλαδή, ένας λανθάνων χώρος (latent space) παρέχει συμπίεση ή υψηλού επιπέδου έννοιες (high-level concepts) των μη επεξεργασμένων δεδομένων που παρατηρούνται, όπως η κατανομή των δεδομένων εισόδου. Στην περίπτωση των GANs, ο γεννήτορας εφαρμόζεται σε σημεία σε ένα επιλεγμένο λανθάνοντα χώρο, έτσι ώστε νέα σημεία που προέρχονται από το χώρο να μπορούν να παρέχονται στον γεννήτορα ως είσοδος και να χρησιμοποιούνται για την παραγωγή νέων και διαφορετικών παραδειγμάτων εξόδου. Μετά την εκπαίδευση, ο γεννήτορας διατηρείται και χρησιμοποιείται για την παραγωγή νέων δειγμάτων.

Ο διευκρινιστής δέχεται ένα παράδειγμα από το πεδίο ορισμού ως είσοδο (πραγματική ή δημιουργημένη) και δίνει ως πρόβλεψη μία ετικέτα δυαδικής κλάσης (‘αληθινή’ ή ‘ψεύτικη’) Το πραγματικό παράδειγμα προέρχεται από το σύνολο δεδομένων εκπαίδευσης. Τα παραγόμενα παραδείγματα εξάγονται από τον γεννήτορα. Ο διευκρινιστής είναι ουσιαστικά ένα μοντέλο ταξινόμησης.

Μετά τη διαδικασία της εκπαίδευσης, ο διευκρινιστής ‘απορρίπτεται’ καθώς εμείς ουσιαστικά ενδιαφερόμαστε για τον γεννήτορα.

Μερικές φορές, ο γεννήτορας μπορεί να χρησιμοποιηθεί όπως έχει μάθει για να εξάγει αποτελεσματικά χαρακτηριστικά από παραδείγματα στο πεδίο του προβλήματος. Μερικά ή όλα τα επίπεδα εξαγωγής χαρακτηριστικών μπορούν να χρησιμοποιηθούν σε εφαρμογές μάθησης μεταφοράς χρησιμοποιώντας τα ίδια ή παρόμοια δεδομένα εισόδου.




Σχήμα 3.3: Δομή ενός GAN (**Πηγή**: developers.google)

### 3.3.4 Εκπαίδευση (Training)

Ο διευκρινιστής συνδέεται με δύο συναρτήσεις απώλειας. Κατά τη διάρκεια της εκπαίδευσής του, ο διευκρινιστής αγνοεί την απώλεια του γεννήτορα και απλά χρησιμοποιεί την απώλεια του διευκρινιστή. Η απώλεια του γεννήτορα χρησιμοποιείται κατά τη διάρκεια της εκπαίδευσής του ιδίου.

Κατά την εκπαίδευσή του, ο διαχωριστής ταξινομεί τόσο τα πραγματικά δεδομένα και όσο τα ψεύτικα δεδομένα που παράγονται από τον γεννήτορα. Η απώλεια του διευκρινιστή 'τιμωρεί' τον διευκρινιστή για την εσφαλμένη ταξινόμηση ενός πραγματικού στιγμιότυπου ως ψεύτικου, ή ψεύτικου ως πραγματικού. Ο διευκρινιστής ενημερώνει τα βάρη του μέσω της ανάστροφης διάδοσης (backpropagation) από την συνάρτηση απώλειάς του μέσω του δικτύου του.

Όσον αφορά την ανάστροφη διάδοση, αρχικά οι τιμές εξόδου κάθε κόμβου υπολογίζονται (και αποθηκεύονται στην κρυφή μνήμη) σε ένα ευθύ πέρασμα. Στη συνέχεια, η μερική παράγωγος του σφάλματος σε σχέση με κάθε παράμετρο υπολογίζεται σε έναν αντίστροφο υπολογισμό μέσω του γραφήματος.

Ο γεννήτορας του GAN μαθαίνει να δημιουργεί ψεύτικα δεδομένα ενσωματώνοντας την ανατροφοδότηση από τον διευκρινιστή. Μαθαίνει να κάνει τον διευκρινιστή να ταξινομεί την έξοδό του ως πραγματική.

Η εκπαίδευση των γεννητόρων απαιτεί στενότερη ολοκλήρωση μεταξύ γεννήτορα και διευκρινιστή από ό,τι απαιτείται στην εκπαίδευση του διευκρινιστή. Το τμήμα του GAN που εκπαιδεύει τον γεννήτορα περιλαμβάνει τυχαία είσοδο, το δίκτυο του γεννήτορα, το οποίο μετασχηματίζει την τυχαία είσοδο σε ένα στιγμιότυπο, το δίκτυο του διευκρινιστή, το οποίο ταξινομεί τα παραγόμενα δεδομένα, την έξοδο του διευκρινιστή και την απώλεια γεννήτορα, η οποία 'τιμωρεί' τον γεννήτορα για την αποτυχία να ξεγελάσει το διευκρινιστή.

Στην πιο βασική μορφή του, ένα GAN λαμβάνει ως είσοδο τυχαίο θόρυβο. Στη συνέχεια, ο γεννήτορας μετασχηματίζει αυτόν το θόρυβο σε έξοδο με νόημα. Με την εισαγωγή του θορύβου, μπορούμε να κάνουμε το GAN να παράγει μια μεγάλη ποικιλία δεδομένων, παίρνοντας δείγματα από διαφορετικά σημεία στην κατανομή-στόχο.

Πειράματα δείχνουν ότι η κατανομή του θορύβου δεν έχει μεγάλη σημασία, έτσι μπορούμε να επιλέξουμε κάτι που μας διευκολύνει στη δειγματοληψία, όπως μια ομοιόμορφη κατανομή. Για λόγους ευκολίας, ο χώρος από τον οποίο γίνεται η δειγματοληψία του θορύβου έχει συνήθως μικρότερη διάσταση από τη διάσταση του χώρου εξόδου.

Για να εκπαιδεύσουμε ένα νευρωνικό δίκτυο, αλλάζουμε τα βάρη του δικτύου ώστε να μειώσουμε το σφάλμα ή την απώλεια της εξόδου του. Στο GAN, ωστόσο, ο γεννήτορας δεν συνδέεται άμεσα με την απώλεια που προσπαθούμε να επηρεάσουμε. Ο γεννήτορας τροφοδοτεί το δίκτυο του διευκρινιστή, και ο διευκρινιστής παράγει την έξοδο που προσπαθούμε να επηρεάσουμε. Η απώλεια γεννήτορα τιμωρεί το γεννήτορα για την παραγωγή ενός δείγματος που το δίκτυο του διευκρινιστή ταξινομεί ως ψεύτικο.

Αυτό το επιπλέον τμήμα του δικτύου πρέπει να περιλαμβάνεται στη διαδικασία ανάστροφης διάδοσης. Η ανάστροφη διάδοση προσαρμόζει κάθε βάρος στη σωστή κατεύθυνση υπολογίζοντας την επίδραση του βάρους στην έξοδο, δηλαδή πώς θα άλλαζε η έξοδος αν άλλαζε το βάρος. Αλλά η επίδραση ενός βάρους του γεννήτορα εξαρτάται από την επίδραση των βαρών του διευκρινιστή. Έτσι η ανάστροφη αναδιάδοση ξεκινά από την έξοδο και ρέει πίσω μέσω του διευκρινιστή στο γεννήτορα.

Την ίδια στιγμή, δεν θέλουμε ο διευκρινιστής να αλλάξει κατά τη διάρκεια της εκπαίδευσης του γεννήτορα. Έτσι εκπαιδεύουμε τον γεννήτορα με την ακόλουθη διαδικασία: Παίρνουμε

δείγμα τυχαίου θορύβου, παράγουμε έξοδο του γεννήτορα από τον τυχαίο δειγματοληπτούμενο θόρυβο, λαμβάνουμε την ταξινόμηση "αληθινό" ή "ψεύτικο" από τον διευκρινιστή για την έξοδο του γεννήτορα, υπολογίζουμε την απώλεια από την ταξινόμηση του διευκρινιστή και κάνουμε ανάστροφη διάδοση μέσω του διευκρινιστή και του γεννήτορα ώστε να γίνει προσαρμογή των βαρών του γεννήτορα.

Αυτή είναι μια επανάληψη της εκπαίδευσης του γεννήτορα. Όσον αφορά την εκπαίδευση του GAN ως σύνολο, αυτή πραγματοποιείται σε εναλλασσόμενες περιόδους. Συγκεκριμένα, ο διευκρινιστής εκπαιδεύεται για μία ή περισσότερες εποχές και έπειτα ο γεννήτορας αντίστοιχα για μία ή περισσότερες εποχές, με αυτή τη διαδικασία να επαναλαμβάνεται.

Διατηρούμε το γεννήτορα σταθερό κατά τη διάρκεια της φάσης εκπαίδευσης του διευκρινιστή. Όσο ο διευκρινιστής προσπαθεί να καταλάβει πως να διακρίνει τα αληθινά δεδομένα από τα ψεύτικα, πρέπει να μάθει πώς να αναγνωρίζει τα ‘ελαττώματα’ του γεννήτορα.

Ομοίως, κρατάμε το διευκρινιστή σταθερό κατά τη διάρκεια της φάσης εκπαίδευσης του γεννήτορα. Είναι αυτή η παλινδρόμηση που επιτρέπει στα GANs να αντιμετωπίσουν τα δυσεπίλυτα παραγωγικά προβλήματα. Ουσιαστικά, αποκτάμε ένα πλεονέκτημα στο δύσκολο παραγωγικό πρόβλημα ξεκινώντας με ένα πολύ απλούστερο πρόβλημα ταξινόμησης. Αντίθετα, αν δεν υφίσταται η εκπαίδευση ενός ταξινομητή για να διακρίνει τη διαφορά μεταξύ των πραγματικών και των παραγόμενων δεδομένων ακόμη και για την αρχική έξοδο της τυχαίας γεννήτριας, δεν μπορεί να ξεκινήσει η εκπαίδευση του GAN.

Καθώς ο γεννήτορας βελτιώνεται με την εκπαίδευση, η απόδοση του διευκρινιστή χειροτερεύει, επειδή ο διευκρινιστής δεν μπορεί εύκολα να εντοπίσει τη διαφορά μεταξύ του πραγματικού και του ψεύτικου. Αν ο γεννήτορας πετύχει, τότε ο διευκρινιστής έχει 50% ακρίβεια, σαν να γυρνά ένα κέρμα για να κάνει την πρόβλεψή του.

Η εξέλιξη αυτή δημιουργεί πρόβλημα για τη σύγκλιση του GAN συνολικά: η ανατροφοδότηση από τον διευκρινιστή αποκτά λιγότερο νόημα με την πάροδο του χρόνου. Αν το GAN συνεχίσει την εκπαίδευση πέρα από το σημείο όπου ο διευκρινιστής δίνει εντελώς τυχαία ανατροφοδότηση, τότε η γεννήτορας αρχίζει να εκπαιδεύεται με λάθος ανατροφοδότηση, και η ποιότητά του μπορεί να χαλάσει. Ουσιαστικά για ένα GAN, η σύγκλιση είναι συχνά μια εφήμερη, παρά μία σταθερή κατάσταση [32, 33].

### 3.3.5 Συναρτήσεις απωλειών (Loss functions)

Τα GAN προσπαθούν να αναπαράγουν μια κατανομή πιθανοτήτων. Επομένως, θα πρέπει να χρησιμοποιούν συναρτήσεις απώλειας που αντικατοπτρίζουν την απόσταση μεταξύ της κατανομής των δεδομένων που παράγονται από το GAN και της κατανομής των πραγματικών δεδομένων.

Δύο συχνές συναρτήσεις απωλειών ειναι:

- η minimax, η οποία χρησιμοποιήθηκε στην αρχική δημοσίευση που εισήγαγε τα GANs
- η Wasserstein, η οποία περιγράφηκε σε δημοσίευση το 2017 και χρησιμοποιείται συχνά στις πρακτικές υλοποιήσεις

Ένα GAN μπορεί να έχει δύο συναρτήσεις απωλειών: μία για την εκπαίδευση γεννητόρων και μια για την εκπαίδευση διευκρινιστών. Ουσιαστικά, στα συστήματα που εξετάζουμε, οι απώλειες γεννητόρων και διευκρινιστών προκύπτουν από ένα μόνο μέτρο της απόστασης μεταξύ των κατανομών πιθανότητας. Και στα δύο αυτά συστήματα, όμως, ο γεννήτορας μπορεί να επηρεάσει μόνο έναν όρο στο μέτρο απόστασης: τον όρο που αντανακλά την κατανομή των πλαστών δεδομένων. Έτσι, κατά τη διάρκεια της εκπαίδευσης των γεννητόρων απορρίπτουμε τον άλλο όρο, ο οποίος αντανακλά την κατανομή των πραγματικών δεδομένων. Στο τέλος, οι απώλειες του γεννήτορα και του διευκρινιστή φαίνονται διαφορετικές, παρόλο που προέρχονται από έναν μόνο τύπο.

Όσον αφορά την συνάρτηση minimax, ο γεννήτορας προσπαθεί να την ελαχιστοποιήσει και ο διευκρινιστής να τη μεγιστοποιήσει:

$$E_x\,[log(D(x))]\ +\ E_z\,[log(1\ -\ D(G(z)))]$$

Σε αυτή τη συνάρτηση:

- $D(x)$, η εκτίμηση του διευκρινιστή για την πιθανότητα ότι το πραγματικό στιγμιότυπο δεδομένων x είναι πραγματικό.
- $E_x$ , η αναμενόμενη τιμή για όλα τα στιγμιότυπα πραγματικών δεδομένων.
- $G(z)$, η έξοδος του γεννήτορα όταν δίνεται θόρυβος z.
- $D(G(z))$, είναι η εκτίμηση του διευκρινιστή για την πιθανότητα ότι ένα ψεύτικο στιγμιότυπο είναι πραγματικό.
- $E_z$ , είναι η αναμενόμενη τιμή πάνω από όλες τις τυχαίες εισόδους στον γεννήτορα (στην πραγματικότητα, η αναμενόμενη τιμή πάνω σε όλα τα παραγόμενα ψεύτικα στιγμιότυπα G(z)).
- Ο τύπος προέρχεται από τη διασταυρούμενη εντροπία (cross-entropy) μεταξύ των πραγματικών και των δημιουργούμενων κατανομών.

Ο γεννήτορας δεν μπορεί να επηρεάσει άμεσα τον όρο $log(D(x))$ στη συνάρτηση, έτσι για τον γεννήτορα, η ελαχιστοποίηση της απώλειας είναι ισοδύναμη με την ελαχιστοποίηση του $log(D(G(z))$

Όσον αφορά την απώλεια Wasserstein, βασίζεται σε μια τροποποίηση του GAN (που ονομάζεται "Wasserstein GAN" ή "WGAN") στο οποίο ο διευκρινιστής δεν ταξινομεί πραγματικά τα στιγμιότυπα. Για κάθε περίπτωση εξάγει έναν αριθμό. Αυτός ο αριθμός δεν χρειάζεται να είναι μικρότερος από 1 ή μεγαλύτερος από 0, επομένως δεν μπορούμε να χρησιμοποιήσουμε το 0.5 ως κατώφλι για να αποφασίσουμε αν ένα στιγμιότυπο είναι πραγματικό ή ψεύτικο. Η εκπαίδευση του διευκρινιστή προσπαθεί απλά να κάνει το αποτέλεσμα μεγαλύτερο για τα πραγματικά στιγμιότυπα από ότι για τα ψεύτικα.

Επειδή δεν μπορεί πραγματικά να διακρίνει μεταξύ αληθινών και ψεύτικων, ο WGAN διευκρινιστής ονομάζεται στην πραγματικότητα "κριτικός" αντί για "διευκρινιστής". Αυτή η διάκριση έχει θεωρητική σημασία, αλλά για πρακτικούς σκοπούς μπορούμε να την αντιμετωπίσουμε ως μια αναγνώριση ότι οι είσοδοι στις συναρτήσεις απωλειών δεν χρειάζεται να είναι πιθανότητες.

Οι συναρτήσεις απωλειών είναι σχετικά απλές:

Κριτική απώλεια:

$$D(x) \; - \; D(G(z))$$

Ο διευκρινιστής προσπαθεί να μεγιστοποιήσει αυτήν τη λειτουργία. Δηλαδή, προσπαθεί να μεγιστοποιήσει τη διαφορά μεταξύ της εξόδου του σε πραγματικές περιπτώσεις και της εξόδου του σε πλαστά στιγμιότυπα.

Απώλεια γεννήτορα:

$$D(G(z))$$

Ο γεννήτορας προσπαθεί να μεγιστοποιήσει αυτήν τη λειτουργία. Δηλαδή, προσπαθεί να μεγιστοποιήσει το αποτέλεσμα του διευκρινιστή για τα ψεύτικα στιγμιότυπά του.

Σε αυτές τις συναρτήσεις:

- $D(x)$ είναι η έξοδος του κριτικού για ένα πραγματικό στιγμιότυπο.
- $G(z)$ είναι η έξοδος του γεννήτορα όταν δίνεται θόρυβος z.
- $D(G(z))$ είναι η έξοδος του κριτικού για ένα ψεύτικο στιγμιότυπο.
- Η έξοδος του κριτικού D δεν χρειάζεται να είναι μεταξύ 1 και 0.
- Οι τύποι προέρχονται από την απόσταση μετακίνησης της γης μεταξύ των πραγματικών και των δημιουργούμενων κατανομών.

Η θεωρητική αιτιολόγηση για το WGAN απαιτεί τα βάρη σε όλο το GAN να θέτονται έτσι ώστε να παραμένουν εντός περιορισμένου εύρους.

Τα Wasserstein GANs είναι λιγότερο ευάλωτα στο να κολλήσουν από τα minimax-βασισμένα GANs, και στο να αποφύγουν τα προβλήματα με τις ολισθαίνουσες κλίσεις. Η απόσταση μετακίνησης της γης έχει επίσης το πλεονέκτημα να είναι μια πραγματική μετρική: ένα μέτρο της απόστασης σε έναν χώρο κατανομών πιθανοτήτων. Αντιθέτως, η διασταυρούμενη εντροπία δεν είναι μετρική με αυτήν την έννοια [34, 35, 36].

## 3.4 Μετάφραση Εικόνας-σε-Εικόνα (Image-to-Image Translation)

### 3.4.1 Ανάλυση

Η μετάφραση εικόνας σε εικόνα έχει ως στόχο τη μεταφορά εικόνων από έναν τομέα προέλευσης σε έναν τομέα προορισμού, διατηρώντας το βασικό περιεχόμενο της αρχικής εικόνας.

Υπάρχουν πολλές εφαρμογές της μετάφρασης εικόνων όπως π.χ. η δημιουργία αληθινής εικόνας από ακμές, ή η αλλαγή της ‘ετικέτας’ σε μία εικόνα.


Σχήμα 3.4: Παράδειγμα μετάφρασης εικόνας-σε-εικόνα (**Πηγή**: https://junyanz.github.io/CycleGAN/)

Είναι ένα δύσκολο πρόβλημα που συνήθως απαιτεί την ανάπτυξη ενός εξειδικευμένου μοντέλου και συνάρτησης απωλειών για το είδος της εργασίας μετάφρασης που εκτελείται.

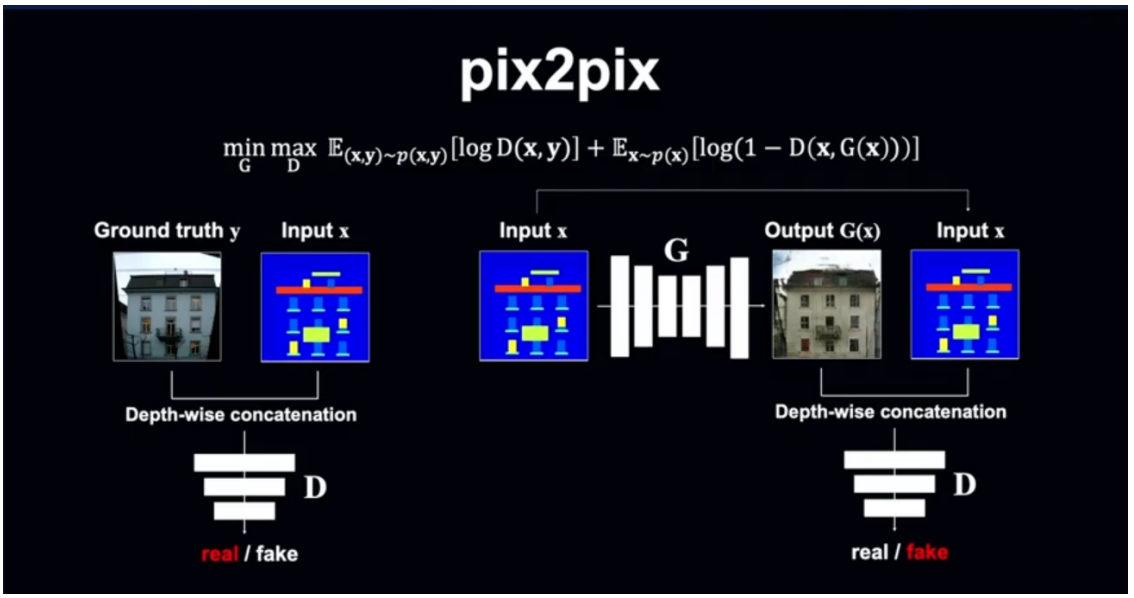
Οι κλασικές προσεγγίσεις χρησιμοποιούν μοντέλα ταξινόμησης ή παλινδρόμησης (regression) ανά-pixel, όπου το πρόβλημα που δημιουργείται είναι ότι κάθε προβλεπόμενο pixel είναι ανεξάρτητο από τα pixels που προβλέπονταν πριν από αυτό και μπορεί να χαθεί η ευρύτερη δομή της εικόνας.

Στην ιδανική περίπτωση, απαιτείται μια τεχνική που είναι γενική, που σημαίνει ότι το ίδιο γενικό μοντέλο και η συνάρτηση απωλειών μπορούν να χρησιμοποιηθούν για πολλές διαφορετικές εργασίες μετάφρασης εικόνας-σε-εικόνα [37, 38].

### 3.4.2 Μετάφραση με ζεύγη: pix2pix

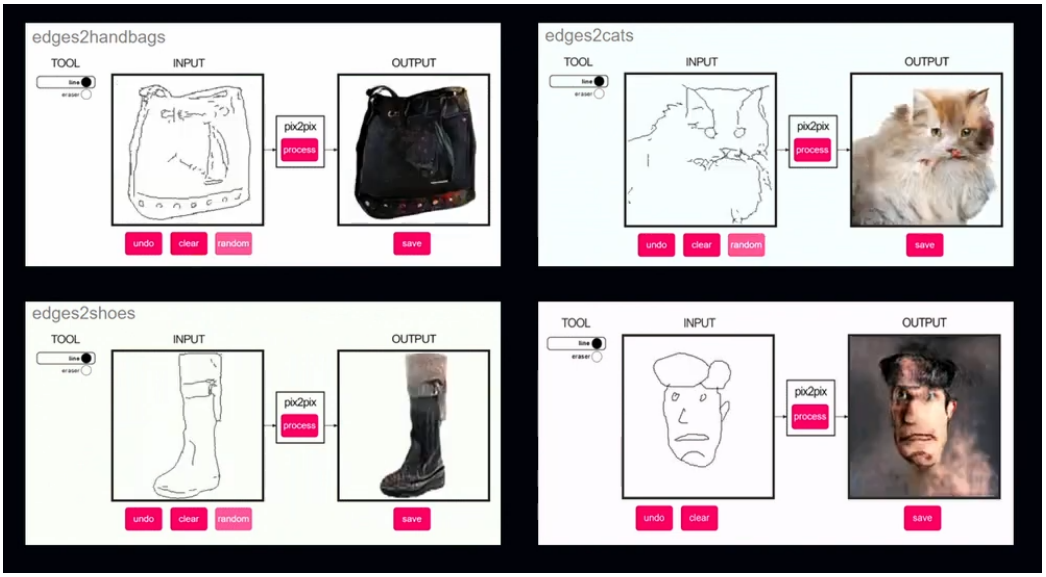
Η πρώτη τεχνική για μετάφραση εικόνων είναι το μοντέλο pix2pix [39]. Το pix2pix βασίζεται στα GANs και ‘μαθαίνει’ δύο μοντέλα ταυτόχρονα, έναν διευκρινιστή και έναν γεννήτορα.

Πρώτα ο διευκρινιστής μαθαίνει να καθορίζει τα αληθινά και ψεύτικα ζεύγη λαμβάνοντας μία εικόνα εισόδου και την αντίστοιχη εικόνα αναφοράς (ground truth image).
Όσον αφορά τον γεννητορα, σκοπός του είναι να ‘κοροϊδέψει’ τον διευκρινιστή με την παραγόμενη έξοδο να είναι όσο το δυνατόν πιο κοντά στην αντίστοιχη σωστή εικόνα.




Σχήμα 3.5: Μοντέλο pix2pix (**Πηγή**: Image-to-Image Translation with Conditional Adversarial Networks)

Στη συνέχεια βλέπουμε κάποια αποτελέσματα μετάφρασης εικόνων με χρήση του pix2pix. Σαν είσοδο έχουμε μία εικόνα ακμών και ως έξοδος δημιουργείται η πραγματική εικόνα.



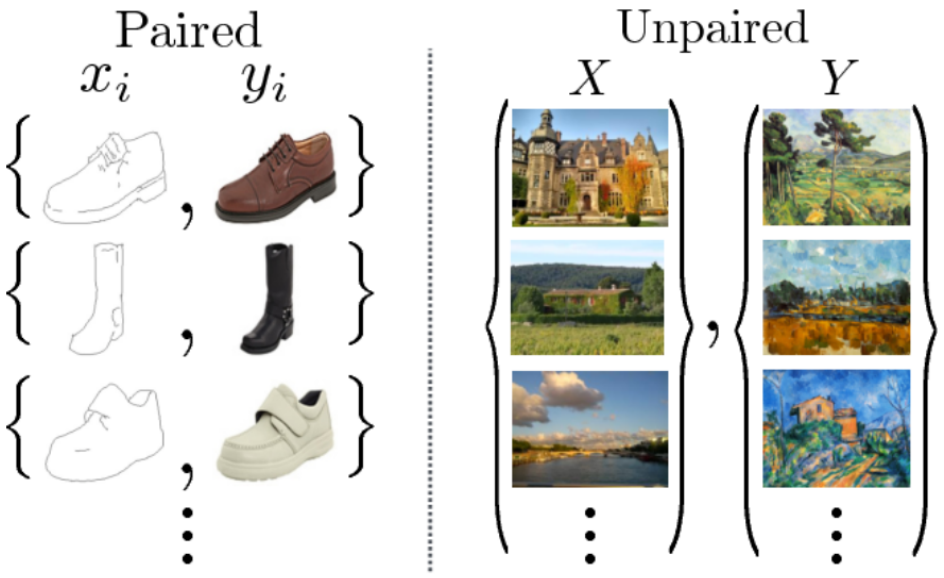

Σχήμα 3.6: Αποτελέσματα pix2pix (**Πηγή**: Image-to-Image Translation with Conditional Adversarial Networks)

Σε συνέχεια του pix2pix αναπτύχθηκε και το pix2pixHD, ένα προχωρημένο μοντέλο το οποίο, όπως υποδηλώνει και η ονομασία του μπορεί να εφαρμοστεί σε υψηλής ανάλυσης εικόνες. Αυτό επιτυγχάνεται με τη προσθήκη κάποιων επιπλέον blocks μπροστά και πίσω από τον generator στο ήδη υπάρχον μοντέλο του pix2pix.

### 3.4.3 Μετάφραση χωρίς ζεύγη: CycleGAN

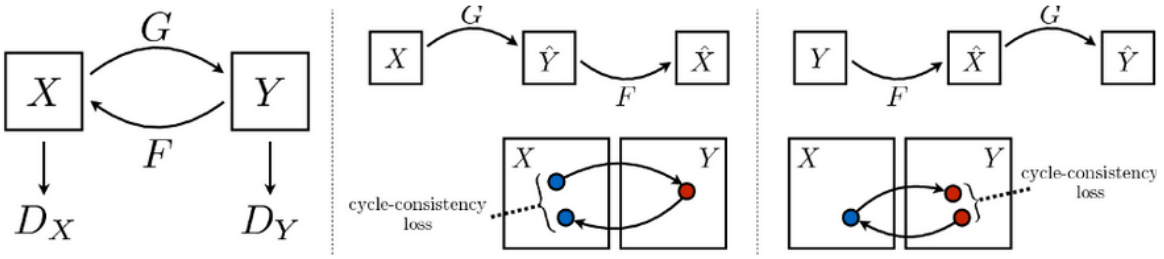
Τα προηγούμενα μοντέλα λειτουργούσαν λαμβάνοντας ζεύγη με είσοδο και έξοδο ως δεδομένα εκπαίδευσης. Ωστόσο η συγκέντρωση τέτοιων δεδομένων σε ζευγάρια είναι αρκετά δύσκολη. Για παράδειγμα, εάν η εργασία της μετάφρασης ήταν η μετατροπή εικόνων που απεικονίζουν γάτες σε εικόνες που απεικονίζουν σκύλους, θα ήταν αρκετά δύσκολο, έως αδύνατο, να βρεθεί ικανός αριθμός εικόνων με γάτες που να έχουν αντιστοιχία με κάποια εικόνα σκύλου, ώστε να δημιουργηθούν τα κατάλληλα ζεύγη.

Για επίλυση αυτού του προβλήματος, υπήρξε η ανάγκη ανάπτυξης κάποιου μοντέλου για μετάφραση εικόνων που να λειτουργεί χωρίς να χρειάζεται ζεύγη για είσοδο, αλλά μόνο ετικετοποίηση κλάσεων.




Σχήμα 3.7: Παράδειγμα μεταφράσεων με και χωρίς ζεύγη (**Πηγή**: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks)

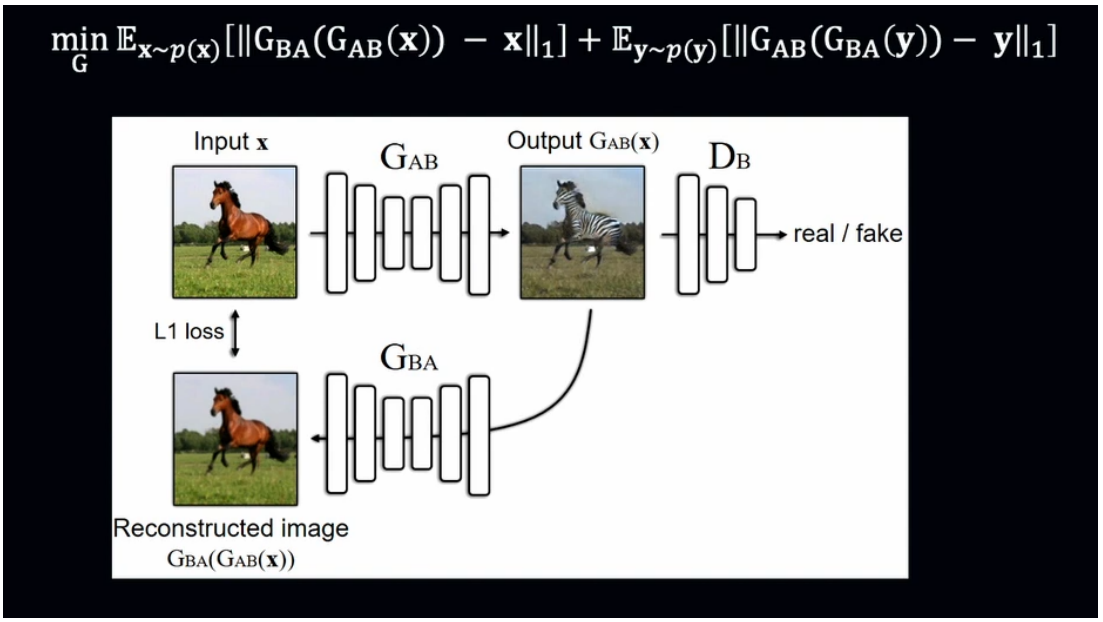
Ένα μοντέλο για μετάφραση εικόνων χωρίς τη χρήση ζευγών είναι το CycleGAN [40], μια προσέγγιση για την εκμάθηση της μετάφρασης μιας εικόνας από έναν τομέα αναφοράς X σε έναν τομέα προορισμού Y. Ο στόχος είναι η εκμάθηση μιας αντιστοίχισης G : X → Y, τέτοιας ώστε η κατανομή των εικόνων από την G(X) να είναι δυσδιάκριτη από την κατανομή Y χρησιμοποιώντας μια ανταγωνιστική απώλεια (adversarial loss). Επειδή αυτή η αντιστοίχιση είναι πολύ περιορισμένη, τη συνδέουμε με μια αντίστροφη αντιστοίχιση F : Y → X και εισαγωγή απώλειας συνέπειας κύκλου (cycle consistency loss) για την επιβολή F(G(X)) ≈ X (και αντίστροφα).




Σχήμα 3.8: Συναρτήσεις και απώλειες του CycleGAN (**Πηγή**: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks)

Το μοντέλο, όπως αναφέρθηκε, περιέχει δύο συναρτήσεις χαρτογράφησης G : X → Y και F : Y → X, και συναφείς ανταγωνιστικούς διακριτές $D_Y$ και $D_X$ . Η $D_Y$ ενθαρρύνει τη G να μεταφράζει το X σε εξόδους που δεν μπορούν να διακριθούν από τον τομέα Y , και αντίστροφα για το $D_X$, το F, και το X . Για την περαιτέρω κανονικοποίηση των αντιστοιχίσεων, εισάγονται δύο απώλειες συνέπειας κύκλου που συλλαμβάνουν τη διαίσθηση ότι αν μεταφράζουμε από τον ένα τομέα στον άλλο και πάλι πίσω θα πρέπει να φτάσουμε εκεί που ξεκινήσαμε.

Παρακάτω φαίνεται το μοντέλο του CycleGAN για μετατροπή εικόνων που αναπαριστούν άλογα σε εικόνες με ζέβρες.



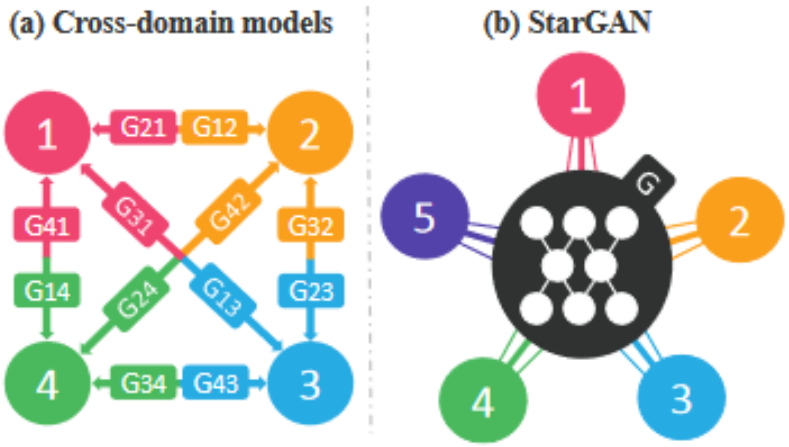

Σχήμα 3.9: Παράδειγμα μετάφρασης εικόνας με CycleGAN (**Πηγή**: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks)

Στο συγκεκριμένο παράδειγμα λαμβάνεται μία εικόνα ως είσοδος. Το CycleGAN εκπαιδεύει τον γεννήτορα να ανακτά την εικόνα εισόδου από την παραγόμενη ώστε να μπορέσει να εκτελέσει την μετάφραση χωρίς να αμελεί τις πληροφορίες της εισόδου. Για παράδειγμα στην παραπάνω μετάφραση, αλλάζοντας μόνο τη μορφή του αλόγου, μπορούμε να κρατήσουμε τη πληροφορία του φόντου, ή εάν πρόκειται για βίντεο, τις κινήσεις του αλόγου.

### 3.4.5 Μετάφραση εικόνων με πολλαπλούς τομείς: StarGAN

Οι προηγούμενες προσεγγίσεις μετάφρασης εικόνων σε εικόνες που περιγράφηκαν έχουν περιορισμένη προσαρμοστικότητα και ανθεκτικότητα στο χειρισμό περισσότερων από δύο τομέων (domains), δεδομένου ότι τα διαφορετικά μοντέλα πρέπει να χτιστούν και να εκπαιδευτούν ανεξάρτητα για κάθε ζεύγος τομέων. Το StarGAN είναι μια νέα και κλιμακούμενη προσέγγιση που μπορεί να εκτελέσει μεταφράσεις εικόνας για πολλαπλούς τομείς χρησιμοποιώντας μόνο ένα μοντέλο [41].

Το StarGan χρησιμοποιεί μόνο ένα ζευγάρι διευκρινιστή-γεννήτορα και μπορεί να μαθαίνει ταυτόχρονα από πολλές βάσεις με διαφορετικά χαρακτηριστικά, δηλαδή μπορεί να 'μεταφράζει' ταυτόχρονα χρώμα μαλλιών και εκφράσεις προσώπου από δύο διαφορετικά σύνολα δεδομένων.



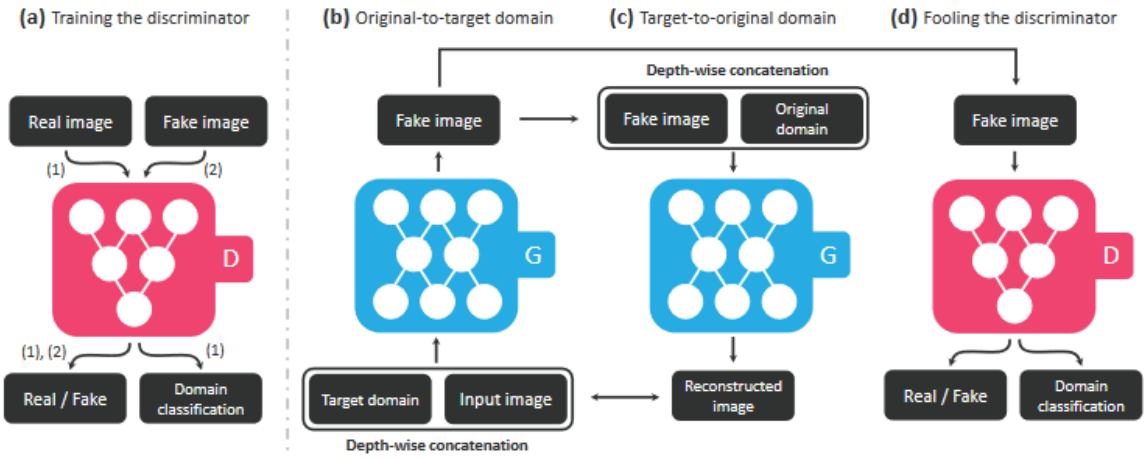

Σχήμα 3.10: Σύγκριση μεταξύ μοντέλου πολλαπλών τομέων και StarGAN. (α) Για το χειρισμό πολλών τομέων, θα πρέπει να δημιουργηθεί από ένα μοντέλο για κάθε ζεύγος τομέων. (β) Το StarGAN είναι ικανό να μαθαίνει αντιστοιχίσεις μεταξύ πολλαπλών τομέων χρησιμοποιώντας ένα μόνο γεννήτορα. Η εικόνα αναπαριστά μια τοπολογία αστέρα που συνδέει πολλούς τομείς. (**Πηγή**: StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation)

Υπάρχουν τρεις βασικές διαδικασίες που συνθέτουν τον αλγόριθμο: ο γεννήτορας, ο διευκρινιστής και ο ανασκευαστής (reconstructor).
Αρχικά, όσον αφορά τον γεννήτορα, έστω ότι έχουμε μία εικόνα με ένα θλιμμένο πρόσωπο που θα μετατραπεί σε εύθυμο. Τροφοδοτούμε την εικόνα στο νευρωνικό δίκτυο του γεννήτορα ο οποίος την μετατρέπει σε μικρότερη ανάλυση. Τα pixels της μικρότερης κλίμακας εικόνας λειτουργούν σαν μεμονωμένοι νευρώνες στο δίκτυο. Στη συνέχεια τα κρυφά στρώματα εντοπίζουν τα χαρακτηριστικά του προσώπου της εικόνας και αλλάζουν κατάλληλα τα pixels που πρέπει για να αλλάξει η έκφραση.

Έπειτα η εικόνα περνάει στον διευκρινιστή ο οποίος πρέπει να αποφανθεί αν είναι αληθινή η ψεύτικη και να κάνει την κατηγοριοποίηση. Δηλαδή εάν ο διευκρινιστής θεωρήσει ότι η εικόνα είναι αληθινή και την κατηγοριοποιήσει ως χαρούμενη έχουμε μία αποδεκτή εικόνα. Σε αντίθετη περίπτωση ο γεννήτορας θα πρέπει να βελτιωθεί.

Η παραγόμενη εικόνα μετατρέπεται πίσω στην αρχική για να διαπιστωθεί πόσο διαφορετική είναι η ανασκευασμένη εικόνα από την πρωτότυπη. Με αυτή τη διαδικασία μπορούμε να προσδιορίσουμε κατά πόσο ο αλγόριθμος έχει παραμορφώσει την πρωτότυπη εικόνα.

Η γενική ιδέα είναι ότι παραγάγουμε μία εικόνα αρκετά αληθινή, ώστε να 'κοροϊδέψει' τον διευκρινιστή, αλλά και να αντιπροσωπεύει τον τομέα-έξοδο που στοχεύουμε αλλά ταυτόχρονα να μην διαφέρει πολύ από την πρωτότυπη εικόνα και άρα να διατηρεί το αρχικό γενικό περιεχόμενο και να μην είναι παραμορφωμένη.




Σχήμα 3.11: Επισκόπηση του StarGAN, που αποτελείται από δύο ενότητες, έναν διευκρινιστή D και έναν γεννήτορα G. (α) ο D μαθαίνει να διακρίνει μεταξύ πραγματικών και ψεύτικων εικόνων και να ταξινομεί τις πραγματικές εικόνες στον αντίστοιχο τομέα. β) Ο G λαμβάνει ως είσοδο τόσο την εικόνα όσο και την ετικέτα του τομέα προορισμού και δημιουργεί μια ψεύτικη εικόνα. Η ετικέτα του τομέα προορισμού αναπαράγεται χωρικά και συνδέεται με την εικόνα εισόδου. (γ) Ο G προσπαθεί να ανακατασκευάσει την αρχική εικόνα από την ψεύτικη με βάση την αρχική ετικέτα τομέα. δ) Ο G προσπαθεί να δημιουργήσει εικόνες που δεν διακρίνονται από τις πραγματικές εικόνες και να ταξινομούνται ως τομέα προορισμού από τον D. (**Πηγή**: StarGAN: Unified Generative Adversarial Networksfor Multi-Domain Image-to-Image Translation)

Όσον αφορά τις συναρτήσεις απωλειών, το StarGAN έχει τρεις βασικές: την ανταγωνιστική απώλεια (adversarial loss), την απώλεια κατηγοριοποίησης τομέα (domain classification loss) και την απώλεια ανακατασκευής (reconstruction loss).

Η ανταγωνιστική απώλεια προκύπτει ως εξής:

$$L_{adv} = E_x [log\, D_{src}(x)] + E_{x,c} [log\, (1 - D_{src}(G(x, c))]$$

Στην συνάρτηση αυτή το $D_{src}(x)$ λαμβάνει τιμές στο διάστημα [0,1] και δείχνει πόσο αληθινό θεωρεί ο διευκρινιστής ότι είναι το $x$, με το 1 να αντιπροσωπεύει το 100% αληθινό και το 0 το 100% ψεύτικο. Ο διευκρινιστής προσπαθεί να μεγιστοποιήσει την συνάρτηση για να επιβεβαιώσει ότι ανιχνεύει μία πραγματική εικόνα $x$ ως αληθινή (δηλαδή το $D_{src}(x)$ να είναι όσο το δυνατόν πιο κοντά στο 1 και το $D_{src}(G(x, c))$, που αντιπροσωπεύει την

εικόνα του γεννήτορα, να είναι όσο το δυνατόν πιο κοντά στο 0). Από την άλλη πλευρά, ο γεννήτορας προσπαθεί να ελαχιστοποιήσει την ίδια συνάρτηση ώστε να πείσει τον διευκρινιστή ότι η ψεύτικη εικόνα είναι αληθινή και ανήκει στην κλάση $c$ ($D_{src}(G(x,c)) \rightarrow 1,\ D_{src}(x) \rightarrow 0$).

Η δεύτερη συνάρτηση είναι η απώλεια κατηγοριοποίησης τομέα:

$$L_{cls}^{r} = E_{x,c'}[-\ log\ D_{cls}(c'|x)] \quad (1)$$

$$L_{cls}^{f} = E_{x,c}[-\ log\ D_{cls}(c|G(x,\ c))] \quad (2)$$

Ο σκοπός του αλγορίθμου δεν είναι μόνο η δημιουργία μιας ρεαλιστικής εικόνας αλλά και η σωστή κατηγοριοποίησή της στον σωστό τομέα $c$. Γι' αυτό ο διευκρινιστής προσπαθεί να ελαχιστοποιήσει την (1) για να επιβεβαιώσει ότι κατηγοριοποιεί μία αληθινή εικόνα $x$ στην κλάση της $c'$. Ο γεννήτορας προσπαθεί να ελαχιστοποιήσει τη (2) για να επιβεβαιώσει ότι ο διευκρινιστής κατηγοριοποιεί την παραγόμενη εικόνα $G$ στον επιθυμητό τομέα $c$.

Η ελαχιστοποίηση των δύο προαναφερθέντων συναρτήσεων απωλειών δεν εγγυάται ότι η 'μεταφρασμένη' εικόνα θα διατηρήσει το αρχικό της περιεχόμενο και δεν θα αλλοιωθεί. Γι' αυτό τον σκοπό εισάγουμε την απώλεια ανακατασκευής:

$$L_{rec} = E_{x,c,c'}[||x - G(G(x,c),\ c')||_1]$$

Εδώ η $G$ παίρνει τη μεταφρασμένη εικόνα $G(x,\ c)$ και την αρχική ετικέτα του τομέα $c'$ ως είσοδο και προσπαθεί να ανακατασκευάσει την αρχική εικόνα $x$. Ιδανικά, η παλιά εικόνα πρέπει να μοιάζει με την ανακατασκευασμένη, οπότε ο γεννήτορας προσπαθεί να ελαχιστοποιήσει αυτή τη συνάρτηση απωλειών.

Παρατίθενται οι ολοκληρωμένες συναρτήσεις απωλειών για τον γεννήτορα και τον διευκρινιστή τις οποίες και οι δύο προσπαθούν να ελαχιστοποιήσουν:

$$L_D = -L_{adv} + \lambda_{cls} L_{cls}^{r}$$

$$L_G = L_{adv} + \lambda_{cls} L_{cls}^{f} + \lambda_{rec} L_{rec}$$

Εδώ τα $\lambda_{cls}$ και $\lambda_{rec}$ είναι υπερ-παράμετροι (hyper-parameters) που ελέγχουν τη σχετική σημασία της ταξινόμησης τομέα και των απωλειών ανακατασκευής, αντίστοιχα, σε σύγκριση με την ανταγωνιστική απώλεια.

Ένα σημαντικό πλεονέκτημα του StarGAN είναι ότι ενσωματώνει ταυτόχρονα πολλαπλά σύνολα δεδομένων που περιέχουν διαφορετικούς τύπους ετικετών, έτσι ώστε το StarGAN να μπορεί να ελέγχει όλες τις ετικέτες στη φάση της δοκιμής (testing). Ένα πρόβλημα όταν πραγματοποιείται εκμάθηση από πολλά σύνολα δεδομένων, όμως, είναι ότι οι πληροφορίες των ετικετών είναι μόνο εν μέρει γνωστές σε κάθε σύνολο. Για παράδειγμα στην περίπτωση των βάσεων δεδομένων CelebA και RaFD, ενώ η πρώτη περιέχει ετικέτες για

χαρακτηριστικά όπως το χρώμα των μαλλιών και το φύλο, δεν διαθέτει ετικέτες για εκφράσεις του προσώπου όπως « χαρούμενος » και « θυμωμένος », και αντιστρόφως για την δεύτερη. Αυτό είναι προβληματικό επειδή οι πλήρεις πληροφορίες για το διάνυσμα ετικετών $c'$ απαιτούνται κατά την ανακατασκευή της εισαγόμενης εικόνας $x$ από τη μεταφρασμένη εικόνα $G(x, c)$.

Για την αντιμετώπιση αυτού του προβλήματος, εισάγουμε ένα διάνυσμα μάσκας (mask vector) m που επιτρέπει στο StarGAN να αγνοήσει τις απροσδιόριστες ετικέτες και να επικεντρωθεί στην ρητά γνωστή ετικέτα που παρέχεται από ένα συγκεκριμένο σύνολο δεδομένων. Στο StarGAN, χρησιμοποιούμε ένα n-διαστάσεων one-hot διάνυσμα για την αναπαράσταση του m, με το n να είναι το πλήθος των συνόλων δεδομένων. Επιπλέον, ορίζουμε μια ενοποιημένη εκδοχή της ετικέτας ως διάνυσμα:

$$\tilde{c} = [c_1, ..., c_n, m]$$

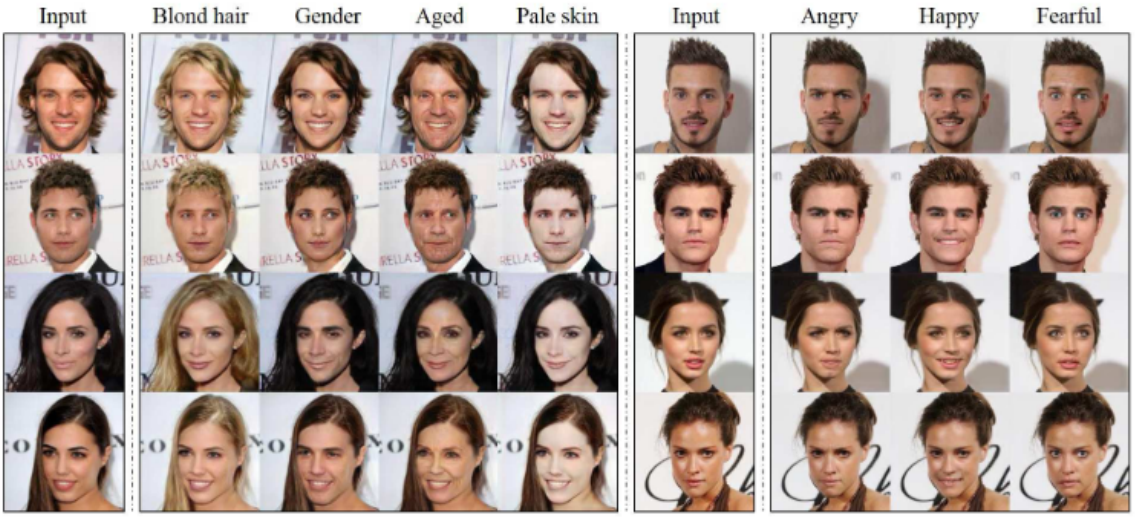
Όσον αφορά την αξιολόγηση της απόδοσης του StarGAN, χρησιμοποιούμε κάποια βασικά μοντέλα για να πραγματοποιήσουμε σύγκριση: το DIAT, το CycleGAN και το IcGAN.

Το DIAT χρησιμοποιεί μια ανταγωνιστική απώλεια για να μάθει την αντιστοίχιση από το x $\in$ X στο y $\in$ Y, όπου το x και το y είναι εικόνες προσώπου σε δύο διαφορετικούς τομείς X και Y, αντίστοιχα. Αυτή η μέθοδος έχει έναν όρο κανονικοποίησης για την αντιστοίχιση ως $||x - F(G(x))||_1$ για τη διατήρηση των χαρακτηριστικών ταυτότητας της εικόνας προέλευσης, όπου F είναι ένα πρόγραμμα εξαγωγής χαρακτηριστικών που έχει προεκπαιδευτεί σε μια εργασία αναγνώρισης προσώπου.

Το CycleGAN, στο οποίο έγινε ήδη αναφορά χρησιμοποιεί επίσης μια ανταγωνιστική απώλεια για να μάθει την αντιστοίχιση μεταξύ δύο διαφορετικών τομέων X και Y . Η μέθοδος αυτή ρυθμίζει τη χαρτογράφηση μέσω απωλειών συνέπειας κύκλου, $||x - (G_{YX}(G_{XY}(x)))||1$ και $||y - (G_{XY}(G_{YX}(y)))||_1$. Αυτή η μέθοδος απαιτεί δύο γεννήτορες και διευκρινιστικές για κάθε ζεύγος δύο διαφορετικών τομέων.

Το IcGAN συνδυάζει έναν κωδικοποιητή με ένα cGAN μοντέλο, το οποίο μαθαίνει την χαρτογράφηση G: {z, c} → x που δημιουργεί μια εικόνα x η οποία εξαρτάται και από το λανθάνον διάνυσμα z και από το υπό συνθήκη διάνυσμα c. Επιπλέον, το IcGAN εισάγει έναν κωδικοποιητή για να μάθει τις αντίστροφες αντιστοιχίσεις του cGAN, $E_Z$ : x → z και $E_c$ : x → c. Αυτό επιτρέπει στο IcGAN να συνθέτει εικόνες αλλάζοντας μόνο το διάνυσμα υπό συνθήκη και διατηρώντας το λανθάνον διάνυσμα.

Τα αποτελέσματα του StarGAN προέκυψαν από το CelebA σύνολο δεδομένων με μεταφορά γνώσης από το RaFD (το οποίο εμπεριέχει εκφράσεις προσώπου) και παρουσιάζονται παρακάτω. Οι περισσότερες είναι αρκετά ρεαλιστικές ωστόσο υπάρχουν ορισμένες ανωμαλίες, για παράδειγμα σε αλλαγή χρώματος στο δέρμα.



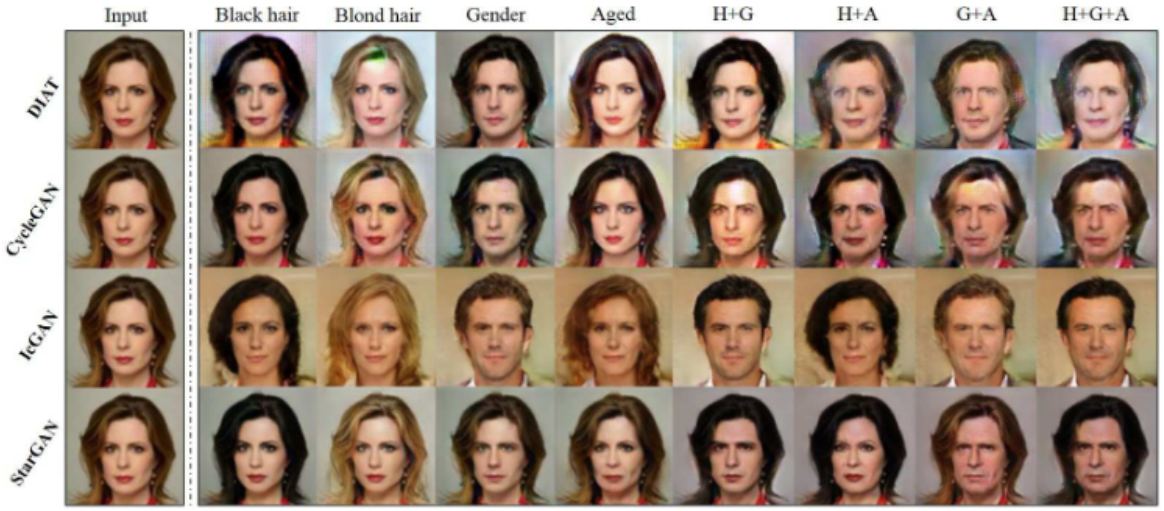

Σχήμα 3.12: Αποτελέσματα StarGAN στο CelebA σύνολο δεδομένων με μεταφορά γνώσης από το RaFD (**Πηγή**: StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation)

Επίσης παρουσιάζεται και σύγκριση του StarGAN με τα βασικά μοντέλα που περιγράψαμε. Για τα πειραματικά αποτελέσματα οι συμμετέχοντες έπρεπε να ψηφίσουν σε ποιά κατηγορία ανήκει η εικόνα και το StarGAN έλαβε τις περισσότερες ψήφους σε όλους τους τομείς, τόσο για έναν όσο και για πολλαπλούς.

| Method | Hair color | Gender | Aged |
|---|---|---|---|
| DIAT | 9.3% | 31.4% | 6.9% |
| CycleGAN | 20.0% | 16.6% | 13.3% |
| IcGAN | 4.5% | 12.9% | 9.2% |
| StarGAN | **66.2%** | **39.1%** | **70.6%** |

| Method | H+G | H+A | G+A | H+G+A |
|---|---|---|---|---|
| DIAT | 20.4% | 15.6% | 18.7% | 15.6% |
| CycleGAN | 14.0% | 12.0% | 11.2% | 11.9% |
| IcGAN | 18.2% | 10.9% | 20.3% | 20.3% |
| StarGAN | **47.4%** | **61.5%** | **49.8%** | **52.2%** |

Σχήμα 3.12: Αντιληπτική αξιολόγηση για την κατάταξη διαφορετικών μοντέλων σε εργασίες μεταφοράς (**Πηγή**: StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation)



Σχήμα 3.13: Αναπαράσταση παραγόμενων εικόνων σε εργασίες μετάφρασης με τη χρήση διαφορετικών μοντέλων (**Πηγή**: StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation)

## 3.5 Συμπεράσματα

Για την εργασία μετάφρασης εικόνας-σε-εικόνα που θα πραγματοποιήσουμε, επιλέγουμε να χρησιμοποιήσουμε το μοντέλο του StarGAN, το οποίο είναι αρκετά σύγχρονο και έχει καταφέρει να αντιμετωπίσει πολλά από τα προβλήματα που είχαν προκύψει στα προγενέστερα μοντέλα όπως pix2pix, συνδυάζοντας επίσης πολλά πλεονεκτήματα και έχοντας πετύχει αρκετά πιο ρεαλιστικά αποτελέσματα συγκριτικά με αντίστοιχα μοντέλα.

# Κεφάλαιο 4. Σύνολα Δεδομένων

## 4.1 Εισαγωγή

Έχοντας καταλήξει στην δομή συναισθημάτων αλλά και στο μοντέλο που θα χρησιμοποιηθεί, αυτό που απομένει είναι η επιλογή ενός κατάλληλου συνόλου δεδομένων (dataset). Το σύνολο αυτό θα χρησιμοποιηθεί για τους σκοπούς της εκπαίδευσης (training) και δοκιμής (testing). Στη συνέχεια θα εξετάσουμε τα διαθέσιμα σύνολα δεδομένων που υπάρχουν για τα επτά βασικά συναισθήματα.

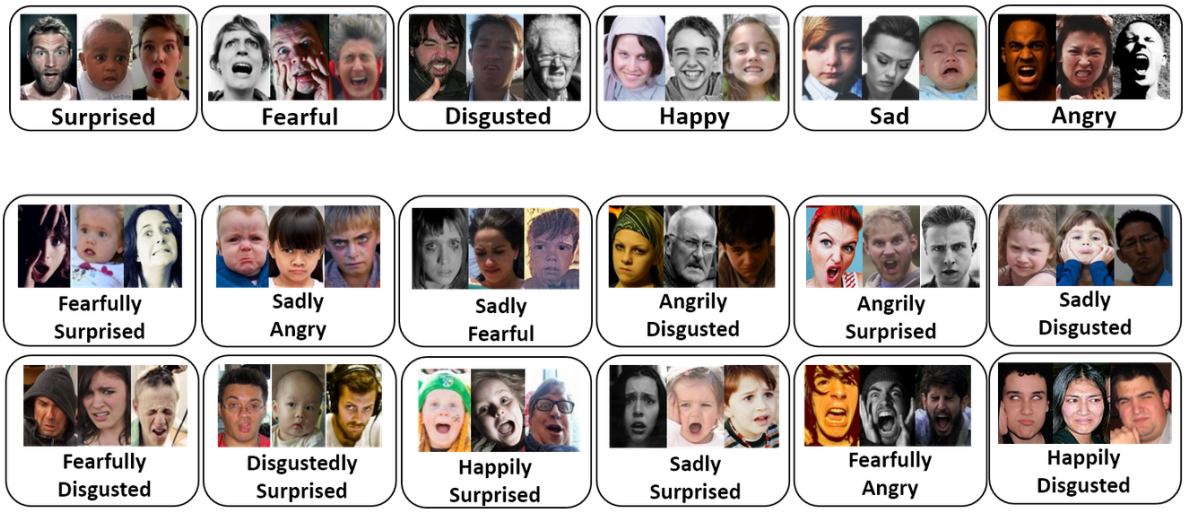
## 4.2 RAF-DB

Η Real-world Affective Faces Database (RAF-DB) [42] είναι μια μεγάλης κλίμακας βάση δεδομένων εκφράσεων προσώπου με περίπου 30 χιλιάδες διαφορετικές εικόνες προσώπου που λαμβάνονται από το Διαδίκτυο. Οι εικόνες σε αυτήν τη βάση δεδομένων παρουσιάζουν μεγάλη μεταβλητότητα ως προς την ηλικία, το φύλο και την εθνικότητα των ατόμων, τις στάσεις του κεφαλιού, τις συνθήκες φωτισμού, τα χαρακτηριστικά και αντικείμενα που εμποδίζουν την ορατότητα του προσώπου (π.χ. γυαλιά και τρίχες του προσώπου), τις μετεπεξεργασίας (π.χ. διάφορα φίλτρα και ειδικά εφέ), κ.λπ.

Η RAF-DB έχει μεγάλη ποικιλία, ποσότητες και σχόλια. Συγκεκριμένα:

- 29672 εικόνες από τον πραγματικό κόσμο,
- Ένα διάνυσμα κατανομής 7 διαστάσεων για κάθε εικόνα,
- Δύο διαφορετικά υποσύνολα: υποσύνολο μίας ετικέτας, συμπεριλαμβανομένων των επτά βασικών συναισθημάτων, υποσύνολο δύο καρτελών, συμπεριλαμβανομένων 12 κατηγοριών κατηγοριών σύνθετων συναισθημάτων,
- 5 ακριβείς τοποθεσίες, 37 αυτόματες τοποθεσίες, πλαίσιο οριοθέτησης, φυλή, εύρος ηλικίας και χαρακτηριστικά φύλου ανά εικόνα,
- Αποτελέσματα ταξινομητή βάσης για βασικά και σύνθετα συναισθήματα.

Για να είναι σε θέση να μετρήσει αντικειμενικά την απόδοση για τις καταχωρήσεις, η βάση δεδομένων έχει χωριστεί σε ένα σύνολο εκπαίδευσης και ένα σύνολο δοκιμών όπου το μέγεθος του συνόλου εκπαίδευσης είναι πέντε φορές μεγαλύτερο από το σύνολο δοκιμής, και οι εκφράσεις και στα δύο σύνολα έχουν σχεδόν ταυτόσημη κατανομή.

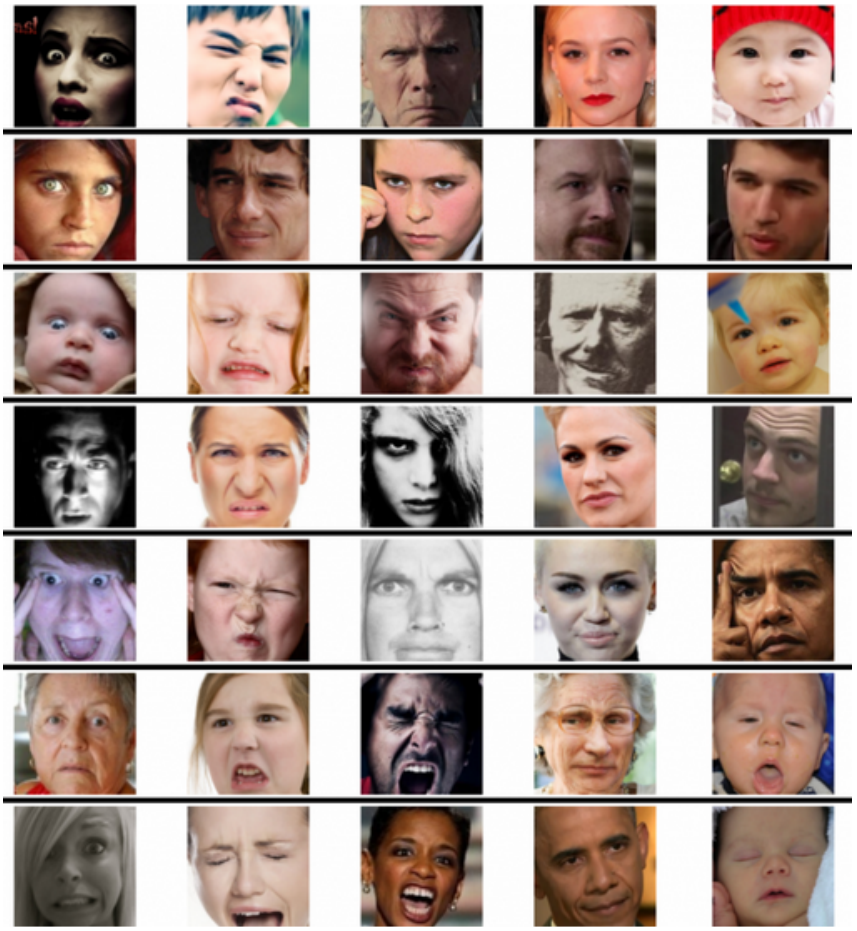



Σχήμα 4.1: Δείγμα εικόνων από την RAF-DB (**Πηγή**: http://www.whdeng.cn/raf/model1.html#dataset)

## 4.3 AffectNet

Οι υπάρχουσες βάσεις δεδομένων με εκφράσεις του προσώπου που υπάρχουν διαθέσιμες είναι μικρές και ως επί το πλείστον καλύπτουν διακριτά συναισθήματα. Υπάρχουν πολύ περιορισμένες σχολιασμένες βάσεις δεδομένων προσώπου στο μοντέλο συνεχούς διάστασης (π.χ., σθένος και διέγερση).

Γι' αυτό τον λόγο, δημιουργήθηκε η βάση AffectNet [43, 44], με συλλογή και σχολιασμό εικόνων με πρόσωπα. Η AffectNet περιέχει περισσότερες από 1M εικόνες που συλλέχθηκαν από το Διαδίκτυο με την βοήθεια τριών μεγάλων μηχανών αναζήτησης που χρησιμοποιούν 1250 -σχετικές με το συναίσθημα- λέξεις-κλειδιά σε έξι διαφορετικές γλώσσες. Περίπου οι μισές από τις αποκτηθείσες εικόνες (~440K) σχολιάστηκαν χειροκίνητα για την παρουσία των επτά βασικών εκφράσεων (κατηγορικό μοντέλο) και την ένταση της σθένους και διέγερσης (διαστατικό μοντέλο). Το γεγονός αυτό καθιστά την AffectNet τη μεγαλύτερη βάση δεδομένων των εκφράσεων προσώπου,επιτρέποντας την έρευνα στην αυτοματοποιημένη αναγνώριση εκφράσεων σε δύο διαφορετικά μοντέλα συναισθημάτων.



Σχήμα 4.2: Δείγμα εικόνων από την AffectNet (**Πηγή**: http://mohammadmahoor.com/affectnet/)

Δύο βαθιά νευρωνικά δίκτυα βάσης χρησιμοποιούνται για να ταξινομήσουν τις εικόνες στο κατηγορικό μοντέλο και να προβλέψουν την ένταση της σθένους και της διέγερσης. Διάφορες μετρήσεις αξιολόγησης δείχνουν ότι οι μέθοδοι που χρησιμοποιούνται με χρήση βαθιών νευρωνικών δικτύων μπορούν να αποδώσουν καλύτερα από τις συμβατικές μεθόδους μηχανικής μάθησης και τα έτοιμα συστήματα αναγνώρισης εκφράσεων προσώπου.

| | Down-Sampling | Up-Sampling | Weighted-Loss |
|---|---|---|---|
| ACCs | 0.50 | 0.47 | 0.58 |
| F1s | 0.49 | 0.44 | 0.58 |
| KAPPAs | 0.42 | 0.38 | 0.51 |
| ALPHAs | 0.42 | 0.37 | 0.51 |
| AUCPR | 0.48 | 0.44 | 0.56 |
| AUC | 0.47 | 0.75 | 0.82 |

Σχήμα 4.3: Πειραματικά αποτελέσματα στο σύνολο επικύρωσης (validation set) για την κατηγοριοποίηση σε οκτώ εκφράσεις (**Πηγή**: http://mohammadmahoor.com/affectnet/)

| | Valence | Arousal |
|---|---|---|
| RMSE | 0.37 | 0.41 |
| CORR | 0.66 | 0.54 |
| SAGR | 0.74 | 0.65 |
| CCC | 0.60 | 0.34 |

Σχήμα 4.4: Πειραματικά αποτελέσματα στο σύνολο επικύρωσης (validation set) για την κατηγοριοποίηση βάσει σθένους-διέγερσης (**Πηγή**: http://mohammadmahoor.com/affectnet/)

| | |
|---|---|
| Neutral | 75374 |
| Happy | 134915 |
| Sad | 25959 |
| Surprise | 14590 |
| Fear | 6878 |
| Disgust | 4303 |
| Anger | 25382 |
| Contempt | 4250 |
| None | 33588 |
| Uncertain | 12145 |
| Non-Face | 82915 |
| Total | 420299 |

Σχήμα 4.5: Συνολικός αριθμός των εικόνων με χειρόγραφα σχόλια στα σύνολα εκπαίδευσης και επικύρωσης σε κάθε κατηγορία συναισθημάτων (**Πηγή**: http://mohammadmahoor.com/affectnet/)

Στην συγκεκριμένη βάση δεν διατίθενται προς το παρόν εικόνες για το στάδιο της δοκιμής (testing) αλλά μόνο για την εκπαίδευση και την επικύρωση (training and validation).

## 4.4 AFF-WILD2

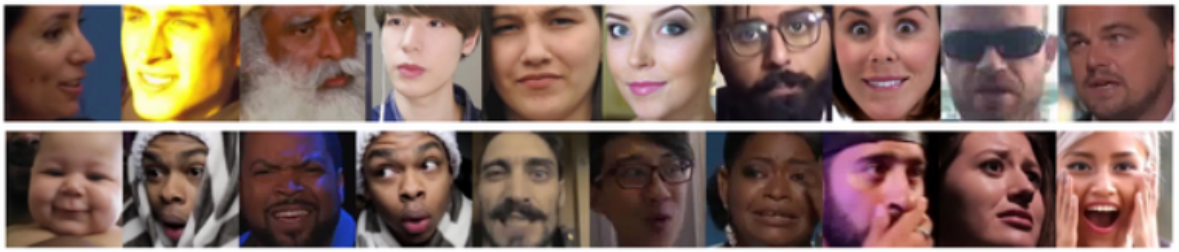
Τα τελευταία χρόνια έχουν προταθεί αρκετές ελεύθερες βάσεις δεδομένων. Ωστόσο οι περισσότερες από αυτές αντιμετωπίζουν ορισμένα από τα εξής προβλήματα:

- το μέγεθός τους είναι μικρό
- δεν έχουν οπτικοακουστικό περιεχόμενο
- μόνο ένα μικρό μέρος είναι χειροκίνητα σχολιασμένο
- περιέχουν μικρό αριθμό θεμάτων

- δεν εμπεριέχουν σχολιασμό για όλες τις κύριες εργασίες συμπεριφοράς (εκτίμηση σθένους-διέγερσης, ανίχνευση μονάδας δράσης και βασική ταξινόμηση έκφρασης).

Για την αντιμετώπιση των παραπάνω αναπτύχθηκε η βάση Aff-Wild, με έμφαση στην μελέτη συνεχών συναισθημάτων, όπως το σθένος και η διέγερση. Η Aff-Wild [45, 46, 47] είναι η πρώτη μεγάλης κλίμακας ελεύθερη βάση δεδομένων και περιέχει περίπου 1.200.000 καρέ. Επιπλέον, ορισμένα τμήματά της επισημαίνονται με παράθεση των βασικών εκφράσεων και τις μονάδες δράσης. Αυτή η βάση επεκτάθηκε με επιπλέον 260 θέματα και 1.413.000 νέα καρέ βίντεο. Η ένωση της Aff-Wild με τα πρόσθετα δεδομένα, ονομάστηκε Aff-Wild2.

Συνολικά, η Aff-Wild2 [48, 49] περιέχει 558 βίντεο με περίπου 2,8 εκατομμύρια καρέ. Εξ όσων γνωρίζουμε, το AffWild2 είναι η πρώτη μεγάλης κλίμακας διαθέσιμη βάση δεδομένων που περιέχει σχόλια και για τις τρεις κύριες δομές συναισθημάτων. Είναι επίσης η πρώτη οπτικοακουστική βάση δεδομένων με σημειώσεις για τις μοναδες δράσης. Όλες οι βάσεις δεδομένων με σχόλια για τις μονάδες δράσης δεν περιέχουν ήχο, αλλά μόνο εικόνες ή βίντεο.



Σχήμα 4.6: Καρέ της Aff-Wild2, που δείχνουν άτομα διαφορετικών εθνικοτήτων, ηλικιακές ομάδες, συναισθηματικές καταστάσεις, πόζες, συνθήκες φωτισμού και αντικείμενα που εμποδίζουν την ορατότητα (**Πηγή**: https://ibug.doc.ic.ac.uk/resources/aff-wild2/)

## 4.5 Συμπεράσματα

Για τους σκοπούς της παρούσας διατριβής επιλέγουμε ως σύνολο δεδομένων την AffectNet. Η AffectNet είναι μία μεγάλη βάση με πολλές ταυτότητες και εύρος, γεγονός που θα βοηθήσει το GAN στην εξαγωγή όσο το δυνατόν καλύτερων αποτελεσμάτων. Γι' αυτό τον λόγο δεν θα χρησιμοποιήσουμε την RAF-DB, καθώς συγκριτικά είναι αρκετά μικρότερης κλίμακας βάση με μικρή ποικιλομορφία.
Όσον αφορά την Aff-Wild2, θα αποτελούσε καλή επιλογή αν στοχεύαμε στην ανάλυση βίντεο. Δεδομένου ότι η ανάλυσή μας επεκτείνεται μόνο στην ανάλυση απλών καρέ (frames), δεν βοηθάει η χρήση μιας βάσης αποτελούμενης από βίντεο.

# Κεφάλαιο 5. Πειραματική Διαδικασία

## 5.1 Εισαγωγή

Έχοντας επιλέξει το κατάλληλο σύνολο δεδομένων για τη δομή συναισθημάτων που θα χρησιμοποιηθεί, μπορούμε να προχωρήσουμε στο πειραματικό στάδιο για το οποίο θα χρησιμοποιήσουμε την υλοποίηση του StarGAN σε tensorflow. Σε αυτό το κεφάλαιο θα αναλυθούν οι απαραίτητες αλλαγές που πραγματοποιήθηκαν και οι παράμετροι που χρησιμοποιούνται. Στη συνέχεια θα γίνει επεξήγηση των σταδίων της εκπαίδευσης και της δοκιμής ώστε να μπορέσουμε στην συνέχεια να προχωρήσουμε στην παρουσίαση των αποτελεσμάτων.

## 5.2 Παράμετροι και υπερ-παράμετροι

### 5.2.1 Εισαγωγή

Οι βασικές παράμετροι ενός νευρωνικού δικτύου είναι τα βάρη των συνδέσεων. Σε αυτήν την περίπτωση, αυτές οι παράμετροι μαθαίνονται κατά τη διάρκεια του σταδίου της εκπαίδευσης. Έτσι, ο ίδιος ο αλγόριθμος, καθώς και τα δεδομένα εισόδου, συντονίζουν αυτές τις παραμέτρους.

Οι υπερ-παράμετροι [50] είναι συνήθως ο ρυθμός εκμάθησης, το μέγεθος παρτίδας και ο αριθμός εποχών. Ονομάζονται ‘υπερ’ επειδή επηρεάζουν τον τρόπο με τον οποίο θα υπολογίζονται οι παράμετροι. Η βελτιστοποίηση αυτών των υπερ-παραμέτρων μπορεί να πραγματοποιηθεί με πολλούς τρόπους: με αναζήτηση πλέγματος (grid search), τυχαία αναζήτηση, με το χέρι, χρησιμοποιώντας απεικονίσεις κ.ο.κ.. Η επιβεβαίωση για το αν οι παράμετροι έχουν υπολογιστεί και για το εάν οι υπερ-παράμετροι είναι αρκετά καλές πραγματοποιείται στο στάδιο της επικύρωσης (validation stage).

### 5.2.2 Εποχές, επαναλήψεις και μέγεθος παρτίδας

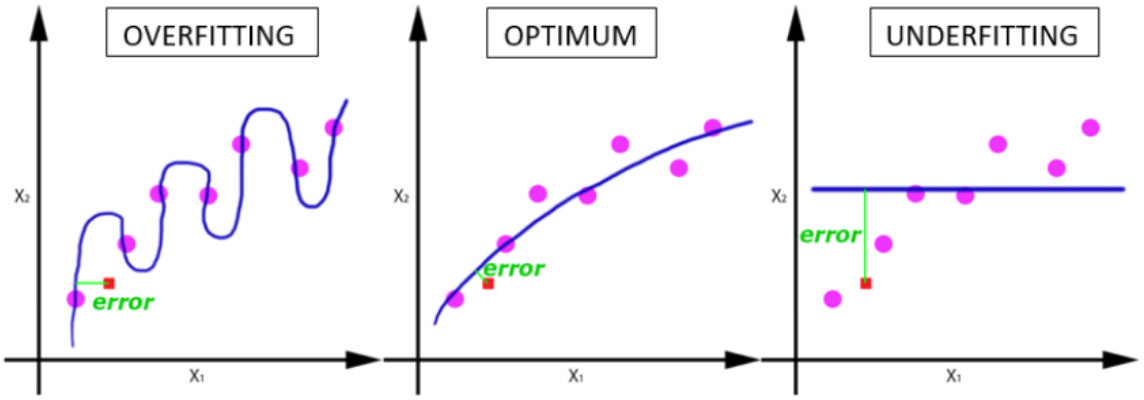
Το νευρωνικό δίκτυο μαθαίνει τα μοτίβα των δεδομένων εισόδου διαβάζοντας το σύνολο δεδομένων εισόδου και εφαρμόζοντας διαφορετικούς υπολογισμούς σε αυτό. Αυτή η διαδικασία μάθησης επαναλαμβάνεται αρκετές φορές χρησιμοποιώντας το σύνολο δεδομένων εισόδου και αντλώντας αποτελέσματα από προηγούμενες δοκιμές. Κάθε διαδρομή για την εκμάθηση από το σύνολο δεδομένων εισόδου ονομάζεται εποχή (epoch).

Ουσιαστικά μια εποχή αναφέρεται σε έναν κύκλο μέσω του συνόλου δεδομένων εκπαίδευσης [51]. Συνήθως, η εκπαίδευση ενός νευρωνικού δικτύου απαιτεί αρκετές εποχές. Η μεγάλη αύξηση του αριθμού των εποχών ωστόσο δεν σημαίνει πάντα ότι το δίκτυο θα έχει καλύτερα αποτελέσματα. Η λάθος επιλογή του αριθμού αυτού μπορεί να οδηγήσει σε υπερεκπαίδευση (overfitting) ή υποεκπαίδευση (underfitting).

Η υπερεκπαίδευση [52] αναφέρεται σε ένα μοντέλο το οποίο εκπαιδεύτηκε υπερβολικά στα στοιχεία της εκπαίδευσης (όταν το μοντέλο δηλαδή μαθαίνει τον θόρυβο στο σύνολο δεδομένων). Ένα τέτοιο μοντέλο δεν αποδίδει καλά σε νέα δεδομένα που δεν έχει. Η υπερεκπαίδευση είναι το πιο συνηθισμένο πρόβλημα στην εφαρμοσμένη μηχανική μάθηση και προκαλει την ψευδαίσθηση ότι ένα μοντέλο είναι εξαιρετικά ακριβές ενώ στην πραγματικότητα δεν αποδίδει καλά.

Αντιθέτως, η υποεκπαίδευση αναφέρεται σε ένα μοντέλο που δεν έχει εκπαιδευτεί επαρκώς. Αυτό μπορεί να οφείλεται σε ανεπαρκή χρόνο εκπαίδευσης ή στο ότι δεν εκπαιδεύτηκε σωστά. Ένα μοντέλο που είναι υποεκπαιδευμένο θα έχει κακές επιδόσεις τόσο στα δεδομένα εκπαίδευσης, όσο και σε νέα.

Τόσο η υπερεκπαίδευση όσο και η υποεκπαίδευση αποφέρουν κακή απόδοση. Η βέλτιστη κατάσταση αναφέρεται ως γενίκευση (generalization). Εδώ το μοντέλο αποδίδει καλά τόσο σε δεδομένα εκπαίδευσης όσο και σε νέα δεδομένα. Έτσι, κάνοντας δοκιμές επιλέγουμε τον αριθμό τον εποχών έτσι ώστε τα αποτελέσματα να παραμένουν τα ίδια μετά από μικρό αριθμό κύκλων [53, 54].




Σχήμα 5.1: Γραφική αναπαράσταση υπερεκπαίδευσης και υποεκπαίδευσης (**Πηγή**: paperspace)

Για κάθε πλήρη εποχή, έχουμε αρκετές επαναλήψεις. Επανάληψη (iteration) είναι ο αριθμός των παρτίδων ή των βημάτων μέσω χωρισμένων πακέτων των δεδομένων εκπαίδευσης που απαιτούνται για την ολοκλήρωση μιας εποχής.

Τέλος, παρτίδα (batch) είναι ο αριθμός των δειγμάτων εκπαίδευσης ή παραδειγμάτων σε μία επανάληψη. Όσο υψηλότερο είναι το μέγεθος παρτίδας, τόσο περισσότερο χώρο μνήμης χρειαζόμαστε.

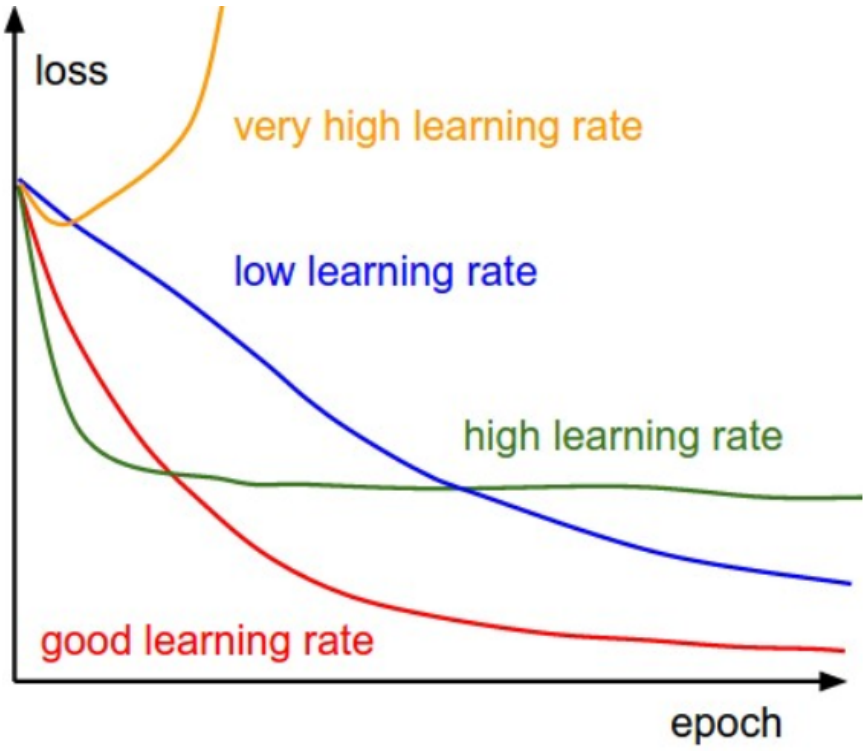
Όσον αφορά τον παρόντα κώδικα, καταλήξαμε στις εξής προεπιλεγμένες υπερ-παραμέτρους:

- αριθμός εποχών: 20
- αριθμός επαναλήψεων: 10,000
- μέγεθος παρτίδας: 16

### 5.2.3 Ρυθμός εκμάθησης

Ο ρυθμός εκμάθησης (learning rate) [55] είναι μια υπερπαράμετρος που ελέγχει κατά πόσο χρειάζεται να αλλάξουμε το μοντέλο μας κάθε φορά που εκπαιδεύουμε τα βάρη του δικτύου. Δεδομένου ότι επηρεάζει σε ποιο βαθμό οι νέες αποκτηθείσες πληροφορίες υπερισχύουν των παλαιών πληροφοριών, αντιπροσωπεύει μεταφορικά την ταχύτητα με την οποία ένα νευρωνικό "μαθαίνει".

Η τιμή του είναι μικρή θετική, συχνά στο εύρος μεταξύ 0.0 και 1.0. Όσο χαμηλότερη είναι η τιμή, τόσο πιο αργά οδεύουμε κατά μήκος της καθοδικής πλαγιάς. Αν και η επιλογή χαμηλού ρυθμού εκμάθησης φαίνεται ιδανική, όσον αφορά τη διασφάλιση ότι δεν χάνουμε τοπικά ελάχιστα, διαλέγοντας υπερβολικά μικρή τιμή ενδεχομένως να επιβραδυνθεί σημαντικά η διαδικασία εκπαίδευσης ή ακόμα και να κολλήσει το μοντέλο σε μια σταθερή κατάσταση. Αντίθετα, αν διαλέξουμε μεγάλη τιμή για τον ρυθμό εκμάθησης τότε είναι πολύ πιθανό να προκληθεί αστάθεια στο σύστημα, δηλαδή ταλαντώσεις ή αδυναμία σύγκλισης στη βέλτιστη λύση [56].




Σχήμα 5.2: Επίδραση διαφόρων ρυθμών μάθησης στη σύγκλιση (convergence) (**Πηγή**: towardsdatascience)

Βάσει των παραπάνω, καταλήξαμε στο να θέσουμε τον ρυθμό εκμάθησης στο 0.0001.

### 5.2.4 Τύπος GAN και ποινή βαθμίδας

Θα γίνει χρήση του Wasserstein GAN με ποινή, ή WGAN-GP, το οποίο χρησιμοποιεί την απώλεια Wasserstein συν μια ποινή-κανόνα βαθμίδας (gradient norm penalty) για την επίτευξη συνέχειας Lipschitz.

Όσον αφορά τη συνέχεια Lipschitz, αυτή ουσιαστικά σημαίνει ότι η συνάρτηση δεν μπορεί να μεταβάλλεται πολύ γρήγορα: η μεταβολή στην τιμή της συνάρτησης, κατ' απόλυτο τιμή, δηλαδή το $|f(x) - f(y)|$ μπορεί να είναι το πολύ ένα πολλαπλάσιο $C|x - y|$ της μεταβολής του ορίσματος κατ' απόλυτο τιμή, δηλαδή του $|x - y|$, με τον συντελεστή C να είναι κοινός για όλα τα ζεύγη x, y [57].

Το αρχικό WGAN χρησιμοποιεί περικοπή βάρους για να επιτύχει τις συναρτήσεις 1-Lipschitz [58], αλλά αυτό μπορεί να οδηγήσει σε ανεπιθύμητη συμπεριφορά με τη δημιουργία παθολογικών επιφανειών (pathological surfaces) και υποχρησιμοποίηση της χωρητικότητας, καθώς και έκρηξη ή εξαφάνιση της βαθμίδας λόγω μη προσεκτικής ρύθμισης της παραμέτρου περικοπής βάρους

Η ποινή βαθμίδας είναι μια απλή εκδοχή του περιορισμού Lipschitz, η οποία προκύπτει από το γεγονός ότι οι συναρτήσεις είναι 1-Lipschitz αν οι βαθμίδες (gradients) είναι το πολύ 1 παντού. Η τετραγωνισμένη διαφορά χρησιμοποιείται ως ποινή (penalty).

Έτσι στον κώδικα θέτουμε:

- gan type: wgan-gp
- gradient penalty lambda: 10

Άλλοι τύποι GAN που θα μπορούσαμε να χρησιμοποιήσουμε είναι το απλό με την κλασσική απώλεια minimax, Hinge, που χρησιμοποιεί διαφορετική συνάρτηση απωλειών [59], ή και WGAN που χρησιμοποιεί διαφορετική ποινή (wgan-lp) [60].

### 5.2.5 Χαρακτηριστικά και εικόνες

Τα χαρακτηριστικά (attributes) που θα χρησιμοποιηθούν σύμφωνα με τις επτά βασικές εκφράσεις είναι: [‘ουδέτερο’, 'θυμός', 'αηδία', 'φόβος', 'χαρά', 'λύπη', 'έκπληξη']
( ['neutral', 'angry', 'disgusted', 'fearful', 'happy', 'sad', 'surprised'] ).

Θεωρούμε ότι σε κάθε εικόνα έχουμε ένα διακριτό κυρίαρχο συναίσθημα και όχι συνδυασμό συναισθημάτων. Αυτό έχει μία απόκλιση από την πραγματικότητα καθώς μπορεί ένα άτομο να βιώνει δύο συναισθήματα ταυτόχρονα (π.χ. χαρά και έκπληξη). Λόγω αυτής της παραδοχής μπορούμε να χρησιμοποιήσουμε ένα one-hot διάνυσμα για τις ετικέτες.

Η one-hot κωδικοποίηση είναι μια μέθοδος μετατροπής δεδομένων ώστε να προετοιμάζονται για αλγόριθμους και να κάνουν καλύτερες προβλέψεις. Με τη μέθοδο one-hot, μετατρέπουμε κάθε κατηγορηματική τιμή σε μια νέα κατηγορηματική στήλη και αντιστοιχίζουμε μια δυαδική τιμή 1 ή 0 στις στήλες. Κάθε ακέραια τιμή αναπαρίσταται ως

ένα δυαδικό διάνυσμα (binary vector). Όλες οι τιμές είναι μηδέν εκτός από τον δείκτη που έχει την τιμή 1 [61].

Έτσι, αν έχουμε τα χαρακτηριστικά που αναφέρθηκαν παραπάνω, για μία εικόνα που απεικονίζει ένα χαρούμενο πρόσωπο θα έχω ετικέτα [0, 0, 0, 0, 1, 0, 0], για μία εικόνα με έκπληξη [0, 0, 0, 0, 0, 0, 1] κ.ο.κ..

Οι εικόνες που χρησιμοποιούνται έχουν μέγεθος 128x128 pixel και είναι έγχρωμες RGB.

Μια εικόνα RGB (red, green, blue — κόκκινο, πράσινο, μπλε) είναι ένας τρισδιάστατος πίνακας byte στον οποίο αποθηκεύεται ρητά μια χρωματική τιμή για κάθε πίξελ. Οι συστοιχίες εικόνων RGB αποτελούνται από πληροφορίες πλάτους, ύψους και τριών καναλιών χρώματος.Οι πληροφορίες χρώματος αποθηκεύονται σε τρία τμήματα μιας τρίτης διάστασης της εικόνας. Αυτές οι ενότητες είναι γνωστές ως κανάλια χρωμάτων, λωρίδες χρωμάτων ή επίπεδα χρωμάτων. Ένα κανάλι αντιπροσωπεύει την ποσότητα του κόκκινου στην εικόνα (το κόκκινο κανάλι), ένα κανάλι αντιπροσωπεύει την ποσότητα του πράσινου στην εικόνα (το πράσινο κανάλι), και ένα κανάλι αντιπροσωπεύει την ποσότητα του μπλε στην εικόνα (το μπλε κανάλι) [62].

Συνεπώς στον κώδικα θα δοθούν οι ακόλουθες τιμές στις αντίστοιχες μεταβλητές:

- image size: 128
- image channels: 3

### 5.2.6 Επαύξηση δεδομένων

Η απόδοση των νευρωνικών δικτύων βαθιάς μάθησης συχνά βελτιώνεται με την ποσότητα των διαθέσιμων δεδομένων.

Η επαύξηση δεδομένων (augmentation) είναι μια τεχνική για την τεχνητή δημιουργία νέων δεδομένων εκπαίδευσης από υπάρχοντα δεδομένα εκπαίδευσης [63]. Αυτό γίνεται με την εφαρμογή ειδικών τεχνικών ανά τομέα σε παραδείγματα από τα δεδομένα εκπαίδευσης που δημιουργούν νέα και διαφορετικά παραδείγματα εκπαίδευσης.

Η επαύξηση των δεδομένων εικόνων είναι ίσως ο πιο γνωστός τύπος αύξησης των δεδομένων και περιλαμβάνει τη δημιουργία μετασχηματισμένων εκδόσεων των εικόνων στο σύνολο δεδομένων εκπαίδευσης, οι οποίες ανήκουν στην ίδια κατηγορία με την αρχική εικόνα.

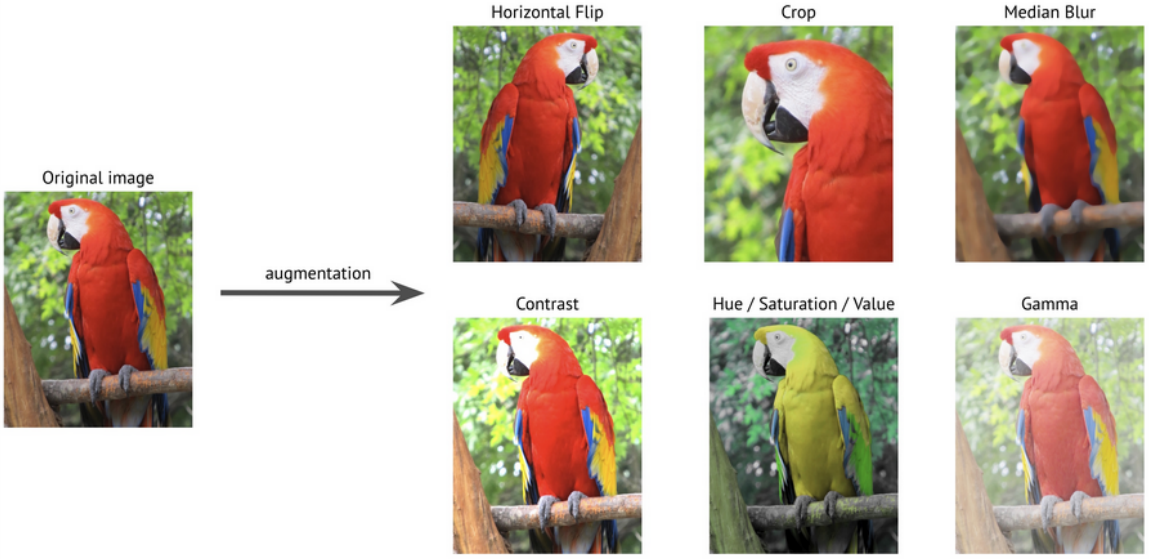
Οι μετασχηματισμοί περιλαμβάνουν ένα εύρος λειτουργιών από το πεδίο του χειρισμού εικόνας, όπως μετατοπίσεις, αντιστροφές, ζουμ κ.α.

Ο στόχος είναι η επέκταση του συνόλου δεδομένων εκπαίδευσης με νέα, πιθανά παραδείγματα. Αυτό σημαίνει, παραλλαγές του συνόλου εκπαίδευσης εικόνες που είναι πιθανό να δει από το μοντέλο. Για παράδειγμα, μια οριζόντια αναστροφή μιας φωτογραφίας μπορεί να έχει νόημα, επειδή η φωτογραφία θα μπορούσε να έχει ληφθεί από τα αριστερά ή τα δεξιά. Αντιθέτως, μια κάθετη αναστροφή π.χ. της φωτογραφίας μιας γάτας δεν έχει νόημα και πιθανότατα δεν θα ήταν κατάλληλη δεδομένου ότι το μοντέλο είναι αρκετά απίθανο να δει μια φωτογραφία ανεστραμμένης γάτας.

Ως εκ τούτου, είναι σαφές ότι η επιλογή των ειδικών τεχνικών αύξησης δεδομένων που χρησιμοποιούνται για ένα σύνολο δεδομένων εκπαίδευσης πρέπει να επιλεγεί προσεκτικά και εντός του πλαισίου του συνόλου δεδομένων εκπαίδευσης και της γνώσης του τομέα του προβλήματος.

Οι σύγχρονοι αλγόριθμοι βαθιάς μάθησης, όπως το συνελικτικό νευρωνικό δίκτυο (CNN), μπορούν να μάθουν χαρακτηριστικά που είναι αναλλοίωτα στη θέση τους στην εικόνα. Παρ' όλα αυτά, η επαύξηση μπορεί να βοηθήσει περαιτέρω σε αυτήν την ανεξάρτητη μετασχηματισμού προσέγγιση για τη μάθηση και μπορεί να βοηθήσει το μοντέλο σε χαρακτηριστικά μάθησης που είναι επίσης αμετάβλητα σε μετασχηματισμούς όπως αριστερά-προς-δεξιά, σε κορυφή-προς-κάτω διάταξη και πολλά άλλα.

Η επαύξηση των δεδομένων των εικόνων συνήθως εφαρμόζεται μόνο στο σύνολο δεδομένων εκπαίδευσης και όχι στο σύνολο δεδομένων επικύρωσης ή δοκιμής. Αυτό διαφέρει από την προεπεξεργασία δεδομένων, όπως η αλλαγή μεγέθους εικόνας και η κλιμάκωση pixels, πρέπει να εκτελούνται με συνέπεια σε όλα τα σύνολα δεδομένων που αλληλεπιδρούν με το μοντέλο.




Σχήμα 5.3: Παράδειγμα επαύξησης δεδομένων (**Πηγή**: https://albumentations.ai/docs/introduction/image_augmentation/)

Έτσι, σαν μεταβλητή στον κώδικα, διατίθεται μία δυαδική 'σημαία' (boolean flag) που μπορεί να τεθεί σε τιμή 'αληθής' ή 'ψευδής', ανάλογα με το εάν χρειάζεται να γίνει επαύξηση δεδομένων ή όχι αντίστοιχα.

## 5.3 Στάδιο εκπαίδευσης

### 5.3.1 Ανάλυση

Γενικά, η εκπαίδευση (training) ενός νευρωνικού δικτύου είναι η διαδικασία εύρεσης των τιμών για τα βάρη και τους σταθερούς όρους. Στις περισσότερες περιπτώσεις, η εκπαίδευση πραγματοποιείται με τη χρήση της τεχνικής εκπαίδευση-δοκιμή. Τα διαθέσιμα δεδομένα, τα οποία έχουν γνωστές τιμές εισόδου και εξόδου, χωρίζονται σε ένα σύνολο εκπαίδευσης (συνήθως το 80% των δεδομένων) και ένα σύνολο δοκιμών (το υπόλοιπο 20%).

Το σύνολο δεδομένων εκπαίδευσης χρησιμοποιείται για την εκπαίδευση του νευρωνικού δικτύου. Ελέγχονται διάφορες τιμές των συντελεστών για τα βάρη και τους σταθερούς όρους για να βρεθεί το σύνολο των τιμών ώστε οι υπολογισμένες τιμές εξόδου να ταιριάζουν περισσότερο με τις  σωστές τιμές. Ουσιαστικά η εκπαίδευση είναι η διαδικασία εύρεσης τιμών για τα βάρη και τους σταθερούς όρους ώστε να ελαχιστοποιείται το σφάλμα. Υπάρχουν πολλοί αλγόριθμοι εκπαίδευσης όπως η ανάστροφη διάδοση.

Κατά τη διάρκεια της εκπαίδευσης, τα δεδομένα της δοκιμής δεν χρησιμοποιούνται καθόλου [64].

### 5.3.2 Εφαρμογή

Το σύνολο δεδομένων εκπαίδευσης που χρησιμοποιούμε από την βάση AffectNet περιλαμβάνει 280 χιλιάδες εικόνες. Η εκπαίδευση πραγματοποιείται σε 20 εποχές, αποτελούμενες από 10 χιλιάδες επαναλήψεις η κάθε μία.

Όσον αφορά την διαδικασία της εκπαίδευσης ενός GAN, όπως αναλύθηκε πραγματοποιείται σε εναλλασσόμενες περιόδους. Συγκεκριμένα, ο διευκρινιστής εκπαιδεύεται για μία ή περισσότερες εποχές και έπειτα ο γεννήτορας αντίστοιχα για μία ή περισσότερες εποχές, με αυτή τη διαδικασία να επαναλαμβάνεται.

Επειδή η διαδικασία της εκπαίδευσης είναι αρκετά χρονοβόρα, γίνεται χρήση ‘σημείων ελέγχου’(checkpoints). Πρόκειται για μια προσέγγιση στην οποία λαμβάνεται ένα στιγμιότυπο της κατάστασης του συστήματος ανά τακτά χρονικά διαστήματα. Έτσι σε περίπτωση κάποιας επιπλοκής θ, υπάρχει αποθηκευμένο το τελευταίο σημείο ελέγχου ώστε η διαδικασία να συνεχίσει από αυτό το σημείο και όχι εξαρχής. Ουσιαστικά το σημείο ελέγχου μπορεί να χρησιμοποιηθεί απευθείας, ή ως σημείο εκκίνησης για μια νέα διαδρομή, συνεχίζοντας από το σημείο που σταμάτησε.

Κατά την εκπαίδευση μοντέλων βαθιάς μάθησης, το σημείο ελέγχου είναι τα βάρη του μοντέλου.

Η συχνότητα με την οποία γίνεται αποθήκευση των σημείων ελέγχου είναι ανά χίλιες επαναλήψεις. Φθάνοντας στο τελευταίο σημείο ελέγχου, υπάρχει η δυνατότητα να το

κρατήσουμε ώστε να χρησιμοποιούμε το μοντέλο ως προεκπαιδευμένο (pretrained), ώστε να μη χρειάζεται να πραγματοποιηθεί επανάληψη της διαδικασίας της εκπαίδευσης.

## 5.4 Στάδιο δοκιμής

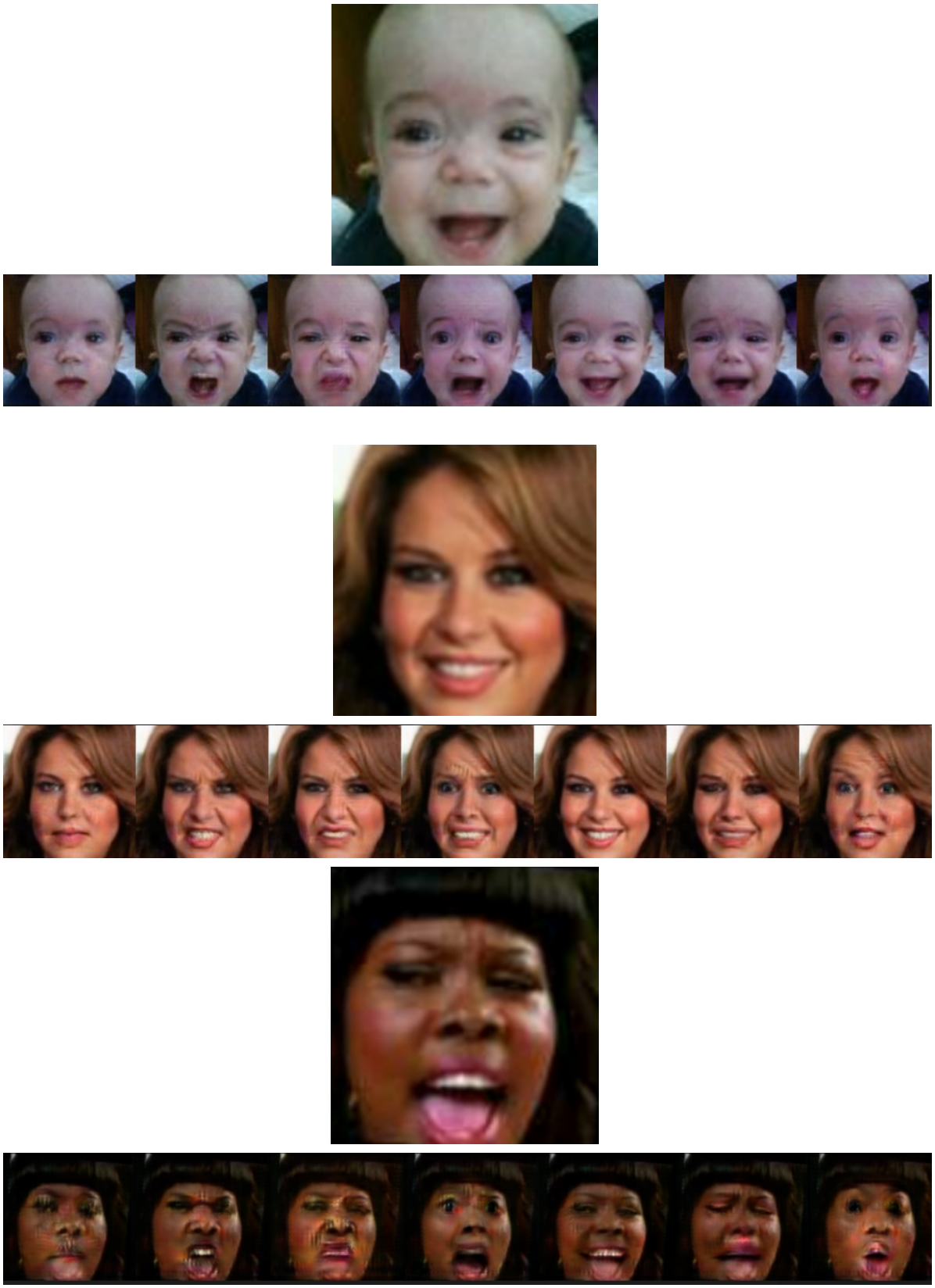
### 5.4.1 Ανάλυση

Αφού ολοκληρωθεί η εκπαίδευση, η ακρίβεια των βαρών και των σταθερών όρων του μοντέλου νευρωνικού δικτύου που προκύπτει εφαρμόζεται μία μόνο φορά στα δεδομένα της δοκιμής (testing). Η ακρίβεια του μοντέλου στα δεδομένα δοκιμής δίνει μια πρόχειρη εκτίμηση του πόσο ακριβές θα είναι το μοντέλο όταν παρουσιάζεται με νέα, προηγουμένως αόρατα δεδομένα, ώστε να επιβεβαιωθεί ότι το μοντέλο εκπαιδεύτηκε επιτυχώς [65].

### 5.4.2 Εφαρμογή

Για το στάδιο της δοκιμής θα χρησιμοποιήσουμε εικόνες που ανήκουν στο σύνολο δεδομένων επικύρωσης της AffectNet καθώς δεν υπάρχουν ακόμη διαθέσιμα δεδομένα για το στάδιο της δοκιμής. Οι εικόνες είναι 5 χιλιάδες στον αριθμό.

Δεδομένου ότι στο συγκεκριμένο στάδιο δοκιμάζουμε τη συμπεριφορά του μοντέλου σε άγνωστα δεδομένα, παίρνουμε και κάποια πρώτα αποτελέσματα παραγόμενων εικόνων τα οποία παρουσιάζονται στη συνέχεια.



Σχήμα 5.4: Πρώτα αποτελέσματα παραγόμενων εικόνων από το στάδιο δοκιμής. Για κάθε πρόσωπο η πάνω εικόνα είναι η πρωτότυπη και οι κάτω οι επτά παραγόμενες. Η σειρά συναισθημάτων είναι η εξής: ουδέτερο πρόσωπο, θυμός, αηδία, φόβος, χαρά, λύπη, έκπληξη.

Αρχικά παρατηρούμε ότι σε κάποιες περιπτώσεις υπάρχει διαφοροποίηση της αρχικής εικόνας με τις τελικές όσον αφορά το χρώμα του προσώπου και γενικά την φωτεινότητα της εικόνας (π.χ. στο δεύτερο παράδειγμα στις παραγόμενες εικόνες έχουμε πιο ροζ τόνους σε σχέση με την αρχική). Αυτό αποτελεί γενικό πρόβλημα του StarGAN και μπορεί να λυθεί με την εισαγωγή μηχανισμών προσοχής (attention mechanisms). Οι μηχανισμοί αυτοί έχουν στόχο τη δημιουργία εικόνων με λεπτομέρειες υψηλής ποιότητας και την απόκτηση συνεπών φόντων σε σχέση με την αρχική εικόνα. Η αλλαγή αυτή στη φωτεινότητα θεωρητικά θα μπορούσε να διαδραματίσει ρόλο στην συμπεριφορά του διευκρινιστή καθώς δεν γνωρίζουμε ακριβώς τι κριτήρια λαμβάνει υπόψιν. Πρακτικά ωστόσο δεν θα έπρεπε να επηρεάσει, ειδικά στην κατηγοριοποίηση του συναισθήματος της εικόνας.

Στην συνέχεια, κάνοντας μία αξιολόγηση των εκφράσεων φαίνεται ότι σε αρκετές περιπτώσεις ορισμένα χαρακτηριστικά είναι αλλοιωμένα. Τα καλύτερα αποτελέσματα για τις περισσότερες περιπτώσεις είναι στην περίπτωση της χαρούμενης έκφρασης και μετά της ουδέτερης, ενώ τα χειρότερα για τις εκφράσεις της αηδίας, του φόβου και της έκπληξης. Αυτό μπορεί να οφείλεται σε αρκετούς παράγοντες. Αρχικά, σίγουρα διαδραματίζει μεγάλο ρόλο η κατανομή των ετικετών. Στις περισσότερες βάσεις, και στην AffectNet, οι εικόνες με τις περισσότερες ετικέτες είναι οι χαρούμενες και ουδέτερες, ενώ με τις λιγότερες, αυτές με έκπληξη, φόβο και αηδία. Ακόμη και στην πραγματική ζωή είναι πιο εύκολο να αναγνωρίσουμε συναισθήματα όπως χαρά και λύπη, ενώ άλλα όπως ο φόβος και η αηδία είναι πιο δύσκολα αναγνωρίσιμα και πολλές φορές παρόμοια μεταξύ τους, ενώ μπορεί να εξαρτώνται και από άλλες παραμέτρους όπως π.χ. η εθνικότητα. Τέλος, όσον αφορά τα συναισθήματα που δεν είχαν τα καλύτερα αποτελέσματα στο στάδιο της δοκιμής, ενδέχεται να ευθύνεται και το γεγονός ότι είναι πιο περίπλοκα προς σύνθεση σε σχέση με αυτά που είχαν καλή απόδοση.

# Κεφάλαιο 6. Αποτελέσματα και Αξιολόγηση

## 6.1 Εισαγωγή

Έχοντας ολοκληρώσει τα στάδια της εκπαίδευσης και δοκιμής, προχωράμε στην αξιολόγηση των δύο μοντέλων μας ξεχωριστά: του διευκρινιστή και του γεννήτορα. Συγκεκριμένα θα χρησιμοποιήσουμε τον γεννήτορα για σύνθεση εικόνων οι οποίες θα αξιολογηθούν για την αληθοφάνειά τους και τον διευκρινιστή, τόσο με εικόνες από το σύνολο δεδομένων όσο και με τις παραγόμενες από τον γεννήτορα και θα ελέγξουμε την ακρίβειά του.

## 6.2 Αξιολόγηση διευκρινιστή με εικόνες από το σύνολο επικύρωσης

Η πρώτη αξιολόγηση του διευκρινιστή θα πραγματοποιηθεί τροφοδοτώντας τον με εικόνες από το σύνολο επικύρωσης. Οι εικόνες αυτές είναι περίπου 5 χιλιάδες. Η διαδικασία αυτή γίνεται στο στάδιο της δοκιμής: Αρχικά φορτώνουμε στο νευρωνικό τα βάρη που έχουν αποθηκευτεί από το στάδιο της εκπαίδευσης. Στη συνέχεια πραγματοποιούμε μία προεπεξεργασία στις εικόνες ώστε να γίνει η κανονικοποίηση τους και τις δίνουμε ως ορίσματα στην συνάρτηση του διευκρινιστή.

Οι έξοδοι που αναμένουμε είναι δύο: η αξιολόγηση της εικόνας προς την αληθοφάνειά της (logit) και το συναίσθημα (c).

Όσον αφορά την πρώτη έξοδο, την λαμβάνουμε ως πίνακα 2x2 και όχι ως μία τιμή (PatchGan [39]). Για να πάρουμε το αποτέλεσμα ως 0 ή 1 (ψεύτικη ή αληθινή εικόνα αντίστοιχα) υπολογίζουμε τη μέση τιμή (mean) κάθε γραμμής του πίνακα και στην συνέχεια κρατάμε την μεγαλύτερη από αυτές τις δύο τιμές. Η στρογγυλοποίησή της μας δίνει την ζητούμενη έξοδο.

Για το συναίσθημα λαμβάνουμε ως έξοδο ένα διάνυσμα επτά θέσεων, όσες και τα συναισθήματα που εξετάζουμε. Εκεί, για την εξαγωγή του αποτελέσματος παίρνουμε το μεγαλύτερο αποτέλεσμα και η θέση στην οποία βρίσκεται είναι το ζητούμενο συναίσθημα (σύμφωνα με τη σειρά που τα έχουμε ορίσει, δηλαδή [‘ουδέτερο’, 'θυμός', 'αηδία', 'φόβος', 'χαρά', 'λύπη', 'έκπληξη'] ).

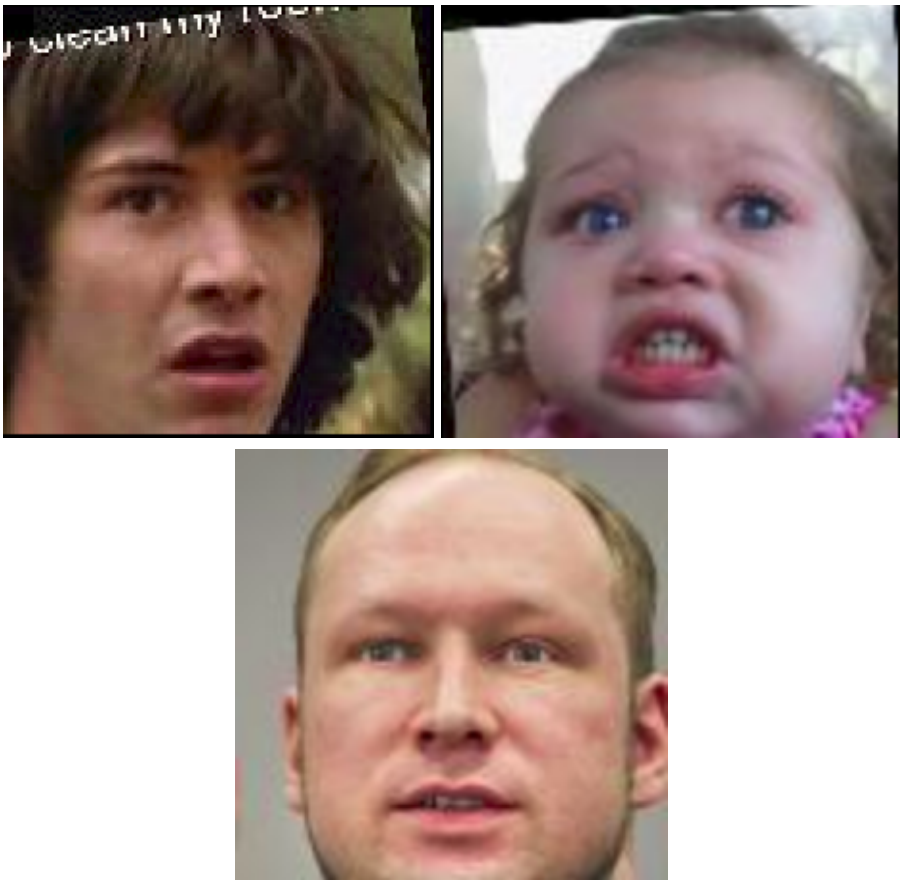
```
[[[ 1.0204773 ]
  [ 3.8433    ]]

 [[ 0.33164018]
  [-1.4066472 ]]]
[ -7.640867  -13.777575  -18.303997  -14.396336    6.786915  -12.2978525
 -10.435304 ]
```

Σχήμα 6.1: Παράδειγμα εξόδων του διευκρινιστή για μία εικόνα: Η πρώτη έξοδος είναι εάν εικόνα είναι αληθινή η ψεύτικη και δίνεται ως πίνακας 2x2. Η δεύτερη είναι το συναίσθημα και δίνεται ως διάνυσμα.

Βγάζουμε αποτελέσματα για περίπου 3 χιλιάδες εικόνες για τις οποίες έχουμε σε αρχείο την αντιστοίχιση της κάθε εικόνας με το σωστό συναίσθημα. Για αυτές τις εικόνες, οι οποίες προφανώς είναι όλες αληθινές, το ποσοστό της επιτυχίας του διευκρινιστή ήταν 97% ως προς την αξιολόγηση για το εάν είναι αληθινές ή ψεύτικες. Για τις αληθινές εικόνες υπήρχε 85% επιτυχία ως προς την αναγνωριση του συναισθήματος, ενώ για τις ψεύτικες εικόνες 100%. Το τελευταίο ποσοστό δεν είναι εντελώς αξιόπιστο καθώς οι εικόνες που ο διευκρινιστής αξιολόγησε ως ψεύτικες ήταν αρκετά λίγες. Συνολικά πάντως η επίδοση του διευκρινιστή είναι αρκετά καλή και για τις δύο εξόδους.

Όσον αφορά την επίδοση του διευκρινιστή ως προς την αξιολόγηση του συναισθήματος παρατηρούμε ότι πραγματοποιείται σύγχυση κυρίως στις εικόνες που απεικονίζουν κυρίως φόβο και αηδία ενώ καλύτερα αποτελέσματα έχουμε για τις χαρούμενες και ουδέτερες. Τα αποτελέσματα αυτά ήταν αναμενόμενα λόγω έλλειψης εικόνων με τα πιο περίπλοκα συναισθήματα (φόβος, αηδία, έκπληξη) στο σύνολο δεδομένων.


Σχήμα 6.2: Εικόνες που ο διευκρινιστής αξιολόγησε εσφαλμένα ως ψεύτικες.

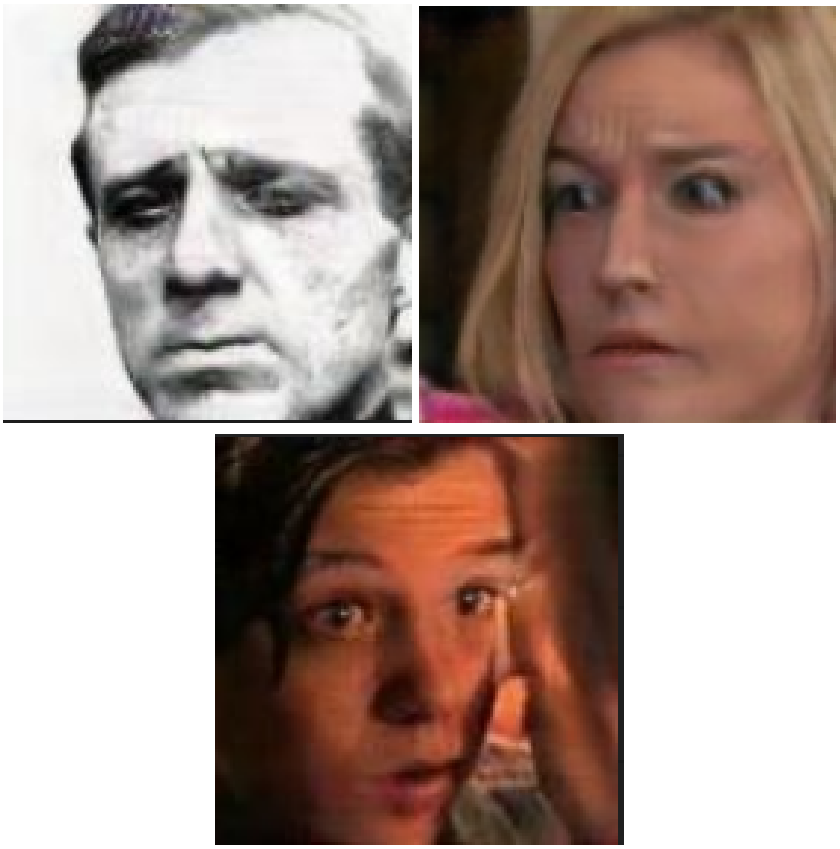


Σχήμα 6.3: Εικόνες που ο διευκρινιστής πραγματοποίησε λάθος κατηγοριοποίηση συναισθήματος: i) Η εικόνα πραγματικά απεικονίζει ‘αηδία’ ενώ ο διευκρινιστής έβγαλε ‘φόβο’ ii) Η εικόνα πραγματικά απεικονίζει ‘φόβο’ ενώ ο διευκρινιστής έβγαλε ‘λύπη’ iii) Η εικόνα πραγματικά απεικονίζει ‘φόβο’ ενώ ο διευκρινιστής έβγαλε ‘ουδέτερο’

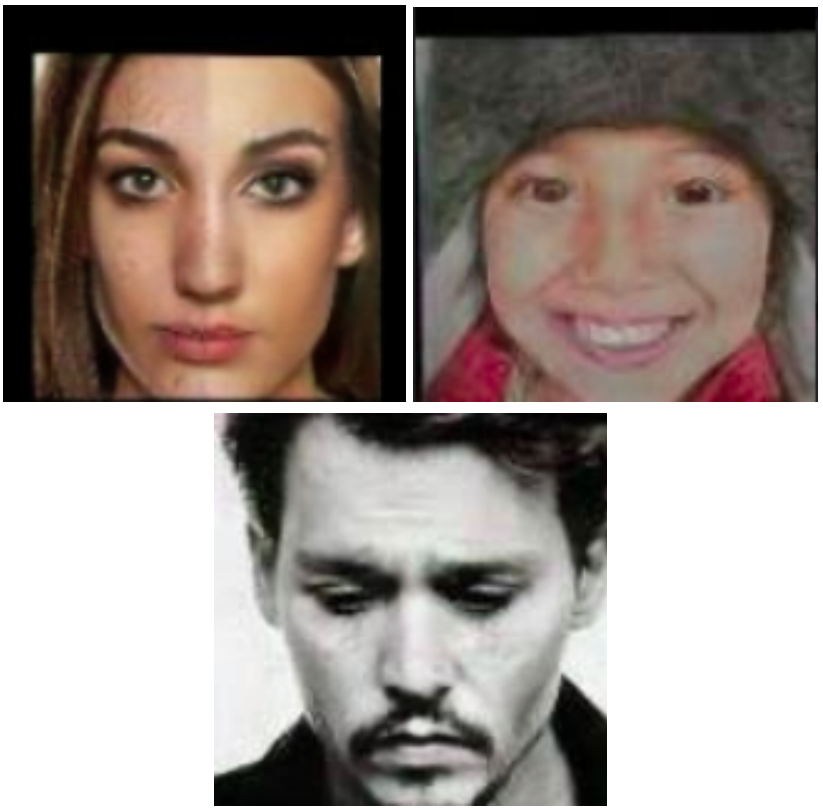
## 6.3 Αξιολόγηση διευκρινιστή με εικόνες από τον γεννήτορα

Τρέχοντας τη συνάρτηση του γεννήτορα για τις εικόνες του συνόλου επικύρωσης και για τη δημιουργία όλων των συναισθημάτων προκύπτουν γύρω στις 35 χιλιάδες εικόνες. Περνώντας τις εικόνες αυτές από την συνάρτηση του διευκρινιστή παίρνουμε τα εξής αποτελέσματα: Ως επί το πλείστον ο γεννήτορας καταφέρει να εξαπατήσει τον διευκρινιστή καθώς το ποσοστό επιτυχίας του διευκρινιστή είναι μόλις 6%, καθώς αξιολόγησε το 94% των εικόνων ως αληθινές ενώ όλες ήταν ψεύτικες. Επίσης όσον αφορά τη σύνθεση των συναισθημάτων, ο γεννήτορας πάλι φαίνεται να απέδωσε πολύ καλά καθώς εδώ έχουμε ακρίβεια 98% στην αξιολόγηση του συναισθήματος. Το ίδιο ισχύει και για τις εικόνες που αναγνωρίστηκαν ως ψεύτικες καθώς σε αυτές υπήρξε ακρίβεια 100%.

Αν και ο γεννήτορας δεν αποτυγχάνει στις περισσότερες περιπτώσεις, παρατηρούμε ότι οι εικόνες  που δεν καταφέρνουν να εξαπατήσουν τον διευκρινιστή, ανήκουν κυρίως στο συναίσθημα του ‘θυμού’. Ακολουθούν οι εκφράσεις της ‘αηδίας’ και της ‘έκπληξης’, πράγμα που και σε αυτή την περίπτωση ήταν αναμενόμενο.



Σχήμα 6.4: Εικόνες γεννήτορα που ο διευκρινιστής αξιολόγησε επιτυχώς ως ψεύτικες.



Σχήμα 6.5: Εικόνες γεννήτορα που ο διευκρινιστής αξιολόγησε ως αληθινές και κατηγοριοποίησε σωστά το συναίσθημα:
i) Ουδέτερη έκφραση ii) Χαρούμενη Έκφραση iii) Λυπημένη έκφραση

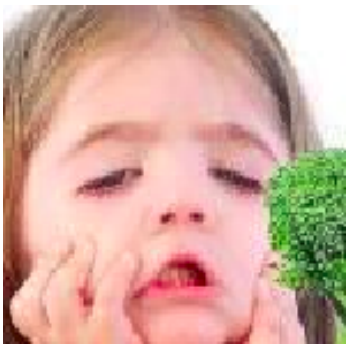
Σχήμα 6.6: Εικόνες γεννήτορα για τις οποίες ο διευκρινιστής πραγματοποίησε λάθος κατηγοριοποίηση συναισθήματος: i) Ο διευκρινιστής κατηγοριοποίησε την έκφραση ως θυμωμένη ενώ ο γεννήτορας την δημιούργησε για λυπημένη ii) Ο διευκρινιστής κατηγοριοποίησε την έκφραση ως χαρούμενη ενώ ο γεννήτορας την δημιούργησε για λυπημένη

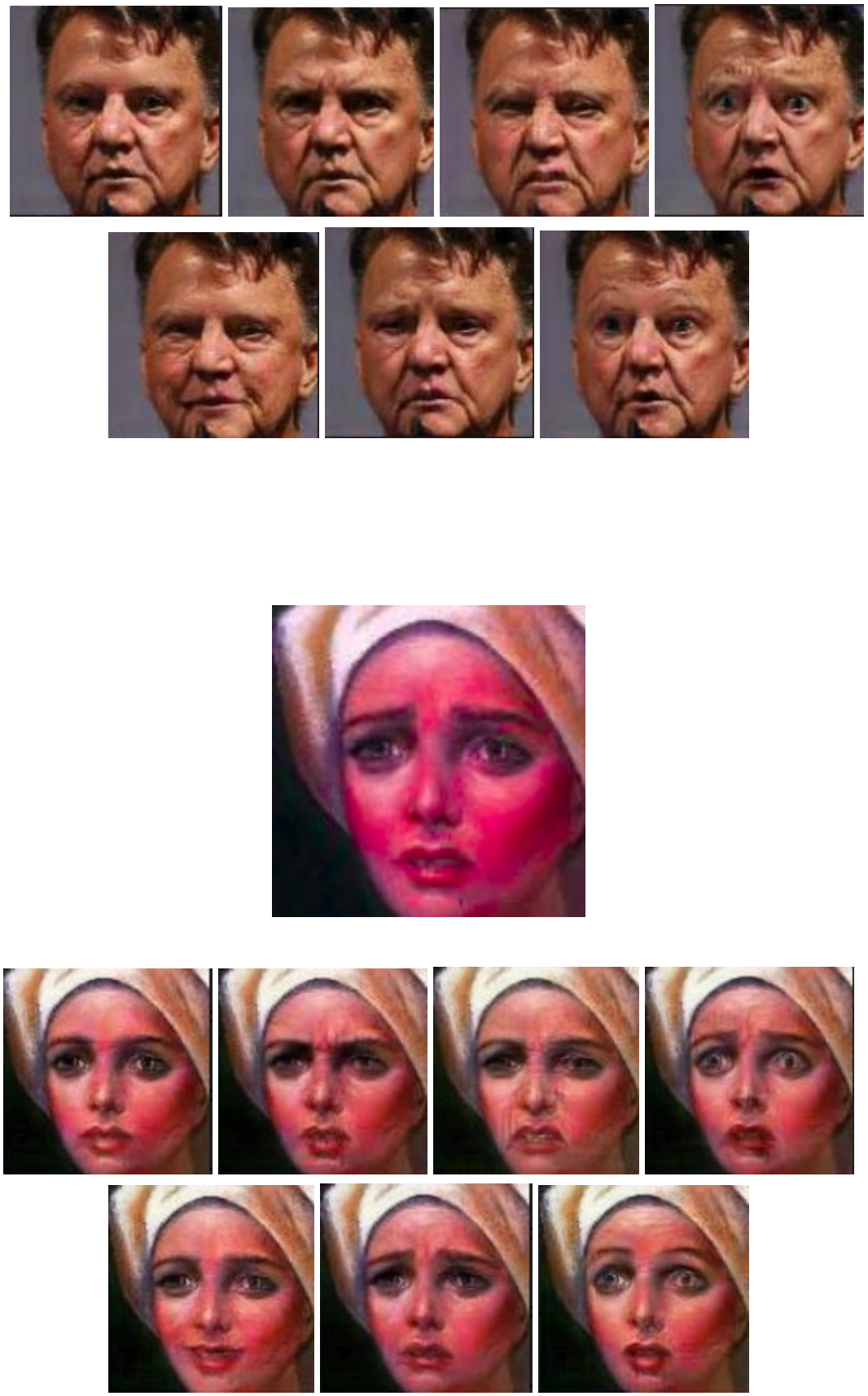
## 6.4 Αξιολόγηση γεννήτορα

### 6.4.1 Αξιολόγηση εικόνων

Αξιολογώντας τις εικόνες που έχει συνθέσει ο γεννήτορας, καταλήγουμε σε παρόμοια συμπεράσματα με εκείνα που είχαμε βγάλει μετά την ολοκλήρωση του σταδίου δοκιμής.

Σε γενικές γραμμές ο γεννήτορας έχει αληθοφανείς εξόδους και κυρίως στο ουδέτερο συναίσθημα και στην χαρά. Λιγότερο καλά αποτελέσματα παρατηρούμε στα πιο σύνθετα συναισθήματα για τα οποία, όπως έχει αναφερθεί, δεν είχαμε και τόσο μεγάλο πλήθος εικόνων για εκπαίδευση. Αυτό έγινε εμφανές και από τα αποτελέσματα του διευκρινιστή. Σε αυτές τις εικόνες πολλές φορές αλλοιώνεται γενικά η μορφή του προσώπου και τα χαρακτηριστικά. Αυτό παρατηρείται επίσης στις εικόνες που το πρόσωπο δεν φαίνεται καθαρά στην αρχική εικόνα (π.χ.. λόγω ύπαρξης αξεσουάρ, όπως γυαλιά). Πολλές φορές σε αυτές τις περιστάσεις ο γεννήτορας δεν κατανοεί την ύπαρξη αυτών ή δεν γνωρίζει πως να τα διαχειριστεί συνεπώς η εικόνα καταλήγει να αλλοιώνεται. Αυτό ευθύνεται πάλι σε μεγάλο βαθμό στο γεγονός ότι το σύνολο δοκιμής δεν περιείχε μεγάλο πλήθος τέτοιων εικόνων. Τέλος, παρατηρούμε ξανά ότι σε αρκετές περιπτώσεις υπάρχει διαφοροποίηση της φωτεινότητας και του γενικού χρώματος της εικόνας. Το γεγονός αυτό ωστόσο δεν φάνηκε να επηρεάζει τον διευκρινιστή, ούτε για την αληθοφάνεια της εικόνας ούτε για την κατηγοριοποίηση του συναισθήματος.





Σχήμα 6.7: Παραδείγματα εικόνων που έχουν δημιουργηθεί από τον γεννήτορα.

### 6.4.2 Μέσο τετραγωνικο σφάλμα

Ένας τρόπος αξιολόγησης της συμπεριφοράς του γεννήτορα είναι με τη χρήση του μέσου τετραγωνικού σφάλματος (mean squared error - MSE).

Όταν δίνουμε ως είσοδο στον γεννήτορα μία εικόνα με συναίσθημα *c* και τον βάλουμε να συνθέσει μία εικόνα με το ίδιο συναίσθημα *c*, τότε οι δύο εικόνες αυτές θα πρέπει να είναι οι ίδιες, ή τουλάχιστον όσο πιο κοντά γίνεται.

Έχοντας συνθέσει τα επτά συναισθήματα για τις εικόνες του συνόλου επιβεβαίωσης, θα υπολογίσουμε το μέσο τετραγωνικό σφάλμα μεταξύ της αρχικής εικόνας και όλων των παραγόμενων. Το μικρότερο μέσο τετραγωνικό σφάλμα από αυτά τα επτά θα υποδηλώνει ότι αυτή η παραγόμενη εικόνα είναι η πλησιέστερη στην αρχική και άρα απεικονίζει το ίδιο συναίσθημα.

Το ελάχιστο τετραγωνικό σφάλμα μεταξύ δύο εικόνων υπολογίζεται ως εξής: Για δύο εικόνες Α, Β λαμβάνουμε το τετράγωνο της διαφοράς μεταξύ κάθε pixel στην εικόνα Α και του αντίστοιχο pixel στην εικόνα Β. Αυτά αθροίζονται και διαιρούνται με τον αριθμό των pixels.

Στο συγκεκριμένο κομμάτι οι αποδόσεις δεν είναι ιδανικές καθώς το μικρότερο τετραγωνικό σφάλμα αντιστοιχίζεται σωστά με την αρχική εικόνα μόνο στο 25% των περιπτώσεων. Αυτό σημαίνει ότι εάν και ο γεννήτορας πραγματοποιεί καλά το κομμάτι της παραγωγής ρεαλιστικών εικόνων και συναισθημάτων, τις περισσότερες φορές αλλοιώνει αρκετά την αρχική εικόνα, με αποτέλεσμα να έχουμε δύο αρκετά διαφορετικές εικόνες για την αναπαράσταση του ίδιου συναισθήματος. Τα καλύτερα αποτελέσματα λαμβάνουμε για τις εικόνες με χαρούμενη έκφραση και ακολουθεί η ουδέτερη.

Σχήμα 6.8: Εικόνες στις οποίες το μικρότερο μέσο τετραγωνικό σφάλμα αντιστοιχούσε στη σωστή αρχική εικόνα: i) Χαρούμενη έκφραση ii) Ουδέτερη έκφραση iii) Λυπημένη έκφραση

Σχήμα 6.9: Εικόνες στις οποίες το μικρότερο μέσο τετραγωνικό σφάλμα δεν αντιστοιχούσε στη σωστή αρχική εικόνα: Η πρώτη στήλη δείχνει την αρχική εικόνα, η δεύτερη την εικόνα με το μικρότερο μέσο τετραγωνικό σφάλμα και η τρίτη την εικόνα που απεικονίζει το ίδιο συναίσθημα με την αρχική
i) Η εικόνα που θα έπρεπε να έχει το μικρότερο τετραγωνικό σφάλμα είναι η λυπημένη και όχι η χαρούμενη ii) Η εικόνα που θα έπρεπε να έχει το μικρότερο τετραγωνικό σφάλμα είναι η έκπληκτη και όχι η φοβισμένη iii) Η εικόνα που θα έπρεπε να έχει το μικρότερο τετραγωνικό σφάλμα είναι η ουδέτερη και όχι η χαρούμενη

# Κεφάλαιο 7. Επίλογος

## 7.1 Γενικά συμπεράσματα

Στην συγκεκριμένη διπλωματική παρουσιάστηκαν τα αποτελέσματα του StarGAN στον τομέα της αναγνώρισης και σύνθεσης συναισθημάτων σε πρόσωπα και εξετάστηκαν οι επιδόσεις του γεννήτορα και του διευκρινιστή.

Ο διευκρινιστής είχε πολύ καλά αποτελέσματα στις αληθινές εικόνες πετυχαίνοντας μεγάλη ακρίβεια τόσο στην δυαδική έξοδο όσο και στην κατηγοριοποίηση του συναισθήματος. Ωστόσο λαμβάνοντας ως είσοδο τις 'ψεύτικες', παραγόμενες από τον γεννήτορα εικόνες, η ακρίβειά του για την κατηγοριοποίηση της εικόνας ως αληθινή/ ψεύτικη, έπεσε κατακόρυφα. Στις περισσότερες περιπτώσεις ο γεννήτορας επιτυγχάνει επιτυχώς τον σκοπό του να εξαπατήσει τον διευκρινιστή, δηλαδή να τον κοροϊδέψει ότι η εικόνα που λαμβάνει είναι αληθινή. Παράλληλα η καλή απόδοση στην κατηγοριοποίηση του συναισθήματος διατηρείται, γεγονός που υποδηλώνει ότι οι παραγόμενες εικόνες δεν είναι μόνο αληθοφανείς αλλά συνθέτουν και επιτυχώς το ζητούμενο συναίσθημα.

Όσον αφορά τον γεννήτορα παρατηρήθηκε ότι, παρά τα φαινομενικά πολύ καλά αποτελέσματα, πραγματοποιεί, στις περισσότερες περιπτώσεις, αλλοίωση της αρχικής εικόνας. Έτσι καταλήγουμε, για το ίδιο συναίσθημα αρχικής και παραγόμενης εικόνας, να μην έχουμε το πλησιέστερο δυνατό αποτέλεσμα. Γενικά, παρατηρούμε ότι σε αρκετές περιπτώσεις υπάρχει διαφοροποίηση της αρχικής εικόνας με τις τελικές όσον αφορά το χρώμα του προσώπου και γενικά την φωτεινότητα της εικόνας. Αρκετά καλά αποτελέσματα παρατηρούμε για τις περιπτώσεις της χαρούμενης έκφρασης και μετά της ουδέτερης, ενώ λιγότερο καλά για τις εκφράσεις της αηδίας, του φόβου και της έκπληξης. Σε αυτό διαδραματίζει τον μεγαλύτερο ρόλο η κατανομή των ετικετών, καθώς οι εικόνες με τις περισσότερες ετικέτες είναι οι χαρούμενες και ουδέτερες, ενώ με τις λιγότερες αυτές με έκπληξη, φόβο και αηδία. Επίσης αξίζει να σημειωθεί ότι γενικά αυτά τα συναισθήματα είναι τα πιο δύσκολο να αναγνωριστούν, α πολλές φορές είναι παρόμοια μεταξύ τους, και μπορεί να εξαρτώνται και από άλλες παραμέτρους. Αυτό το γεγονός τα καθιστά ι πιο περίπλοκα προς σύνθεση σε σχέση με αυτά που είχαν καλή απόδοση. Τα παραπάνω είχαν αντίκτυπο και στον διευκρινιστή καθώς σε αυτά τα πιο 'δύσκολα' συναισθήματα αντιλαμβανόταν πιο εύκολα ότι η εικόνα ήταν ψεύτικη, όμως παράλληλα επειδή ούτε ο ίδιος είχε αρκετές εικόνες για να εκπαιδευτεί έκανε περισσότερα λάθη στην κατηγοριοποίησή τους.

## 7.2 Μελλοντικές Επεκτάσεις

Η παρούσα εργασία μπορεί να επεκταθεί υλοποιώντας κάποια επιπλέον κομμάτια: Αρχικά όσον αφορά την αλλοίωση της αρχικής εικόνας που πραγματοποιεί ο γεννήτορας, αποτελεί γενικό πρόβλημα του StarGAN που μπορεί να ξεπεραστεί με την εισαγωγή μηχανισμών προσοχής (attention mechanisms). Οι μηχανισμοί αυτοί συμβάλλουν στη δημιουργία

εικόνων με λεπτομέρειες υψηλής ποιότητας και την απόκτηση συνεπών φόντων σε σχέση με την αρχική εικόνα. Η προσθήκη αυτή θα βοηθούσε στην επίτευξη καλύτερων αποτελεσμάτων στην αξιολόγηση του γεννήτορα και ενδεχομένως να συνέβαλε στην αποτελεσματικότερη εξαπάτηση του διευκρινιστή, αν και δεν φαίνεται να διαδραμάτισε σημαντικό ρόλο στον συγκεκριμένο τομέα. Αλλη επέκταση θα μπορούσε να είναι η χρήση των ‘ψεύτικων’ εικόνων του γεννήτορα για εκπαίδευση ενός νευρωνικού δικτύου (ενός ResNet) και η δοκιμή αυτού με χρήση αληθινών εικόνων από το σύνολο επικύρωσης. Με αυτόν τον τρόπο θα επιτευχθεί καλύτερη ανάλυση της επίδοσης του γεννήτορα. Τέλος επιπλέον επέκταση μπορεί να γίνει εκμεταλλεύοντας την βασική ιδιότητα του StarGan για μετάφραση πολλαπλών τομέων και να προσθέταμε και άλλον ή άλλους τομείς (domains).

# Βιβλιογραφία


[1] "The Value of Emotion Recognition Technology | IT Business Edge", IT Business Edge, 2021. [Online]. Available: https://www.itbusinessedge.com/business-intelligence/value-emotion-recognition-technology/.

[2] "AI Emotion and Sentiment Analysis With Computer Vision in 2021 - viso.ai", *viso.ai*, 2021. [Online]. Available: https://viso.ai/deep-learning/visual-emotion-ai-recognition/.

[3] Y. Avrithis, N. Tsapatsoulis and S. Kollias, "Broadcast news parsing using visual cues: a robust face detection approach," 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532), 2000, pp. 1469-1472 vol.3, doi: 10.1109/ICME.2000.871044.

[4] N. Tsapatsoulis and S. Kollias, "Face detection in color images and video sequences," 2000 10th Mediterranean Electrotechnical Conference. Information Technology and Electrotechnology for the Mediterranean Countries. Proceedings. MeleCon 2000 (Cat. No.00CH37099), 2000, pp. 498-502 vol.2, doi: 10.1109/MELCON.2000.879979.

[5] L. Malatesta, A. Raouzaiou, K. Karpouzis, and S Kollias. Towards modeling embodied conversational agent character profiles using appraisal theory predictions in expression synthesis. Applied intelligence, 30(1):58–64, 2009.

[6] George Caridakis, Amaryllis Raouzaiou, Kostas Karpouzis, Stefanos Kollias. Synthesizing gesture expressivity based on real sequences. Workshop on multimodal corpora: from multimodal behaviour theories to usable models, LREC 2006 Conference.

[7] D. Kollias, G. Marandianos, A. Raouzaiou and A. Stafylopatis, "Interweaving deep learning and semantic techniques for emotion analysis in human-machine interaction," *2015 10th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, 2015, pp. 1-6, doi: 10.1109/SMAP.2015.7370086.

[8] D. Kollias, A. Tagaris and A. Stafylopatis, "On line emotion detection using retrainable deep neural networks," *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850049

[9] D. Kollias, M. Yu, A. Tagaris, G. Leontidis, A. Stafylopatis and S. Kollias, "Adaptation and contextualization of deep neural network models," *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1-8, doi: 10.1109/SSCI.2017.8280975.

[10] D. Kollias, S. Cheng, E. Ververas, I. Kotsia, and S. Zafeiriou. Deep neural network augmentation: Generating faces for affect analysis. International Journal of Computer Vision, pages 1–30, 2020

[11] D. Kollias and S. Zafeiriou. Va-stargan: Continuous affect generation. In International Conference on Advanced Concepts for Intelligent Vision Systems, pages 227–238. Springer, 2020.

[12] N. Haber, C. Voss and D. Wall, "Upgraded Google Glass Helps Autistic Kids “See” Emotions", *IEEE Spectrum*, 2020. [Online]. Available: https://spectrum.ieee.org/upgraded-google-glass-helps-autistic-kids-see-emotions/particle-1.

[13] E. Digitale, "Google Glass helps kids with autism read facial expressions", *News Center*, 2018. [Online]. Available: https://med.stanford.edu/news/all-news/2018/08/google-glass-helps-kids-with-autism-read-facial-expressions.html.

[14] R Plutchik. Emotion: A psychoevolutionary synthesis harper & row new york. 1980.

[15] K. Cherry, "Scientists Study Core Emotions vs. Those Influenced by Culture", *Verywell Mind*, 2021. [Online]. Available: https://www.verywellmind.com/how-many-emotions-are-there-2795179.

[16] Humintell.com, 2010. [Online]. Available: https://www.humintell.com/2010/06/the-seven-basic-emotions-do-you-know-them/.

[17] D Kollias, A Schulc, E Hajiyev, and S Zafeiriou. Analysing affective behavior in the first abaw 2020 competition. In 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG), pp. 794–800, 2020.

[18] D. Kollias, I. Kotsia, E. Hajiyev, and S. Zafeiriou. Analysing affective behavior in the second abaw2 competition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.3652-3660, 2021.

[19] Y. -I. Tian, T. Kanade and J. F. Cohn, "Recognizing action units for facial expression analysis," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 97-115, Feb. 2001, doi: 10.1109/34.908962.

[20] "Facial Action Coding System - Wikipedia", En.wikipedia.org, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Facial_Action_Coding_System.

[21] JA Russell, A circumplex model of affect, Journal of Personality and Social Psychology, 1980

[22] Posner J, Russell JA, Peterson BS. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol. 2005.

[23] "Deep learning - Wikipedia", En.wikipedia.org, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Deep_learning.

[24] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. S. Awwal, and V. K. Asari, "A State-of-the-Art Survey on Deep Learning Theory and Architectures," Electronics, vol. 8, no. 3, p. 292, Mar. 2019.

[25] D. Ciregan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification," 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3642-3649, doi: 10.1109/CVPR.2012.6248110.

[26] D. Arora, M. Garg and M. Gupta, "Diving deep in Deep Convolutional Neural Network," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 749-751, doi: 10.1109/ICACCCN51052.2020.9362907.

[27] "Generative Adversarial Networks", Google Developers, 2019. [Online]. Available: https://developers.google.com/machine-learning/gan/.

[28] J. Brownlee, "A Gentle Introduction to Generative Adversarial Networks (GANs)", Machine Learning Mastery, 2019. [Online]. Available: https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/

[29] Love, B.C. Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review* **9,** 829–835 (2002). Available: https://doi.org/10.3758/BF03196342

[30] Barlow HB. Unsupervised learning. Neural computation. 1989

[31] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. "Unsupervised learning." The elements of statistical learning. Springer, 2009. 485-585.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In NIPS, 2014.

[33] Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. IEEE Signal Processing Magazine. 2018

[34] Pan Z, Yu W, Wang B, Xie H, Sheng VS, Lei J, Kwong S. Loss functions of generative adversarial networks (GANs): opportunities and challenges. IEEE Transactions on Emerging Topics in Computational Intelligence, 2020.

[35] Frogner C, Zhang C, Mobahi H, Araya-Polo M, Poggio T. Learning with a Wasserstein loss. arXiv preprint arXiv:1506.05439, 2015.

[36] Dukler Y, Li W, Lin A, Montúfar G. Wasserstein of Wasserstein loss for learning generative models. InInternational Conference on Machine Learning, 2019.

[37] Y. Hao, "Image-to-Image Translation", Medium, 2019. [Online]. Available: https://towardsdatascience.com/image-to-image-translation-69c10c18f6ff.

[38] Liu MY, Breuel T, Kautz J. Unsupervised image-to-image translation networks. InAdvances in neural information processing systems,2017 (pp. 700-708).

[39] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1125-1134).

[40] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. InProceedings of the IEEE international conference on computer vision 2017 (pp. 2223-2232).

[41] Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 8789-8797).

[42] S. Li and W. Deng, "Real-world Affective Faces (RAF) Database", Whdeng.cn, 2016. [Online]. Available: http://www.whdeng.cn/raf/model1.html#dataset.

[43] Mollahosseini A, Hasani B, Mahoor MH. Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing. 2017 Aug 21;10(1):18-31.

[44] M. Mahoor, "AffectNet", *Mohammadmahoor.com*, 2017. [Online]. Available: http://mohammadmahoor.com/affectnet/.

[45] Zafeiriou S, Kollias D, Nicolaou MA, Papaioannou A, Zhao G, Kotsia I. Aff-wild: valence and arousal In-the-Wild Challenge. InProceedings of the IEEE conference on computer vision and pattern recognition workshops 2017 (pp. 34-41).

[46] Kollias D, Tzirakis P, Nicolaou MA, Papaioannou A, Zhao G, Schuller B, Kotsia I, Zafeiriou S. Deep affect prediction in-the-wild: Aff-wild database and challenge, deep architectures, and beyond. International Journal of Computer Vision. 2019 Jun 127(6):907-29.

[47] Benroumpi A, Kollias D. AffWild Net and Aff-Wild Database 2019 Oct 11.

[48] Kollias D, Zafeiriou S. Aff-wild2: Extending the aff-wild database for affect recognition. arXiv preprint arXiv:1811.07770. 2018 Nov 11.

[49] Kollias D, Zafeiriou S. Expression, affect, action unit recognition: Aff-wild2, multi-task learning and arcface. arXiv preprint arXiv:1910.04855. 2019 Sep 25.

[50] P. Radhakrishnan, "What are Hyperparameters", 2017. [Online]. Available: https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a.

[51] Brownlee J. What is the Difference Between a Batch and an Epoch in a Neural Network?. Machine Learning Mastery. 2018 Jul 20.

[52] Hawkins DM. The problem of overfitting. Journal of chemical information and computer sciences. 2004 Jan 26;44(1):1-2.

[53] Jabbar H, Khan RZ. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). Computer Science, Communication and Instrumentation Devices. 2015:163-72.

[54] Van der Aalst WM, Rubin V, Verbeek HM, van Dongen BF, Kindler E, Günther CW. Process mining: a two-step approach to balance between underfitting and overfitting. Software & Systems Modeling. 2010 Jan;9(1):87-111.

[55] "Learning rate - Wikipedia", En.wikipedia.org, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Learning_rate.

[56] Li Y, Wei C, Ma T. Towards explaining the regularization effect of initial large learning rate in training neural networks. arXiv preprint arXiv:1907.04595. 2019 Jul 10.

[57] Gouk H, Frank E, Pfahringer B, Cree MJ. Regularisation of neural networks by enforcing lipschitz continuity. Machine Learning. 2021 Feb;110(2):393-416.

[58] Cobzaş Ş, Miculescu R, Nicolae A. Lipschitz functions. Cham: Springer; 2019

[59] Kavalerov I, Czaja W, Chellappa R. A multi-class hinge loss for conditional gans. InProceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision 2021 (pp. 1290-1299).

[60] Zhou C, Zhang J, Liu J. Lp-WGAN: Using Lp-norm normalization to stabilize Wasserstein generative adversarial networks. Knowledge-Based Systems. 2018 Dec 1;161:415-24.

[61] A. Fawcett, "What is One Hot Encoding?", Educative: Interactive Courses for Software Developers, 2021. [Online]. Available: https://www.educative.io/blog/one-hot-encoding.

[62] "Indexed and RGB Image Organization", Climserv.ipsl.polytechnique.fr, 2021. [Online]. Available: https://climserv.ipsl.polytechnique.fr/documentation/idl_help/Indexed_and_RGB_Image_Organization.html.

[63] Wang X, Wang K, Lian S. A survey on face data augmentation for the training of deep neural networks. Neural computing and applications. 2020 Mar 17:1-29.

[64] Larochelle H, Bengio Y, Louradour J, Lamblin P. Exploring strategies for training deep neural networks. Journal of machine learning research. 2009 Jan 1;10(1).

[65] Sun Y, Huang X, Kroening D, Sharp J, Hill M, Ashmore R. Testing deep neural networks. arXiv preprint arXiv:1803.04792. 2018 Mar 10.