



Πανεπιστήμιο Δυτικής Αττικής

Σχολή Μηχανικών

Τμήμα Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής

Διπλωματική Εργασία

**Μηχανική Μάθηση και Ανθρώπινοι Παράγοντες στην
Αυτόνομη Οδήγηση**

Παρασκευόπουλος Ιωάννης

71445047

Επιβλέπων:

Νικολάου Γρηγόριος

Λέκτορας Πανεπιστημίου

Αιγάλεω – Αθήνα, Οκτώβριος, 2021

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Παρασκευόπουλος Ιωάννης του Δημητρίου, με αριθμό μητρώου 71445047 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος ΜΗΧΑΝΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΠΑΡΑΓΩΓΗΣ,

δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.»

Ο Δηλών
Παρασκευόπουλος Ιωάννης



Η Διπλωματική Εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τριμελή επιτροπή:

Γ. Νικολάου, Λέκτορας Πανεπιστημίου	Σ. Βασιλειάδου, Επίκουρη Καθηγήτρια	Χ. Δρόσος, Εργαστηριακό Διδακτικό Προσωπικό (ΕΔΠ)
(Υπογραφή)	(Υπογραφή)	(Υπογραφή)

Copyright © Με επιφύλαξη παντός δικαιώματος. All rights reserved.
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και Παρασκευόπουλος Ιωάννης,
Οκτώβριος, 2021.

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

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Δυτικής Αττικής.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κύριο Νικολάου Γρηγόριο για την υπέροχη συνεργασία καθώς και για την καθοδήγησή του, που με βοήθησε να φέρω εις πέρας την εργασία αυτή.

Τέλος θα ήθελα να ευχαριστήσω και την οικογένειά μου, που ήταν δίπλα μου σε κάθε βήμα και με στήριζαν ένθερμα.

Περίληψη

Μία εκ των βασικών αιτιών για την απώλεια ανθρωπίνων ζωών είναι η απόσπαση της προσοχής των οδηγών, σύμφωνα με διεθνής στατιστικές μελέτες. Για την επίλυση των ανωτέρω προβλημάτων απαιτείται η εκπαίδευση νευρωνικών δικτύων με τη χρήση της βάσης δεδομένων State Farm Distracted Driver Detection και στατιστικών αυτής. Η Βαθιά και Μηχανική Μάθηση – τα δύο υποπεδία της Τεχνητής Νοημοσύνης – παρέχουν τα εργαλεία που χρειάζονται για την εκπαίδευση των υπολογιστών, προκειμένου να αντιμετωπιστούν τα διάφορα προβλήματα.

Έπειτα γίνεται εκτενής αναφορά στην μέθοδο επίλυσης του προβλήματος με τη χρήση νευρωνικών δικτύων. Οι δοκιμές για την αξιολόγηση των επιδόσεων γίνεται πάνω σε δεδομένα με και χωρίς ετικέτες. Ακόμη, πραγματοποιείται και δοκιμή σε δεδομένα σε πραγματικό χρόνο.

Τέλος, παρουσιάζονται τα αποτελέσματα της μελέτης, αλλά και προτάσεις για μελλοντικές βελτιώσεις.

Abstract

One of the leading causes of loss of life is the distraction of drivers, according to international statistical studies. Solving the above problems requires training of neural networks using the State Farm Distracted Driver Detection database and its statistics. Deep Learning and Machine Learning - the two subfields of Artificial Intelligence - provide the tools needed for computer training to tackle various problems.

Extensive reference is then made to the method of solving the problem using neural networks. Performance evaluation tests are performed on data with and without labels. Real-time data testing is also performed.

Finally, the results of the study are presented, as well as suggestions for future improvements.

Περιεχόμενα

Κεφάλαιο 1^ο : Εισαγωγή	8
1.1 Εισαγωγή	8
1.1.1 Παράγοντες Κινδύνου.....	8
1.1.2 Απόσπαση της προσοχής του οδηγού	8
1.1.3 Στατιστικά.....	8
1.1.4 Συμπέρασμα.....	15
1.1.5 Άλλα συστήματα που αυξάνουν την ασφάλεια στην οδήγηση.....	15
1.2 Τα Δεδομένα	17
1.2.1 Το σύνολο δεδομένων “State Farm Distracted Driver Detection”.....	17
1.2.2 Περιγραφή των δεδομένων	17
1.2.3 Στατιστικά.....	18
1.3 Τεχνητή Νοημοσύνη (Artificial Intelligence).....	21
1.3.1 Σύντομη Ιστορική Αναδρομή.....	21
1.3.2 Τι είναι η Τεχνητή Νοημοσύνη (AI);.....	21
1.3.3 Μηχανική Μάθηση	22
1.3.3 Βαθιά Μάθηση.....	23
1.3.4 Πώς λειτουργεί ένα νευρωνικό δίκτυο;.....	25
1.3.5 Τι έχει πετύχει η βαθιά μάθηση μέχρι σήμερα;	28
1.3.6 Βαθιά μάθηση για μηχανική όραση.....	28
1.3.7 Η λειτουργία της συνέλιξης	29
1.3.8 Τα τρία είδη εργασιών της Μηχανικής Όρασης	31
Κεφάλαιο 2: Μέθοδος Επίλυσης	32
2.1 Tensorflow & Keras.....	32
2.2 Γενίκευση και Βελτιστοποίηση	32
2.3 Υπερπροσαρμογής (Overfitting) & Υποπροσαρμογής (Underfitting).....	32
2.3.1 Τρόποι καταπολέμησης της Υπερπροσαρμογής (Overfitting) & Υποπροσαρμογής (Underfitting)	33
2.3.2 Αύξηση δεδομένων (Data Augmentation)	36
2.3.3 Θόλωση προσώπων.....	37
2.3.4 Βάρη Κατηγοριών (Class Weights)	38
2.3.5 Τμηματοποίηση Εικόνας (Image Segmentation)	39

2.4 Σχετικές Εργασίες.....	43
2.5 Η μεθοδολογία.....	45
2.5.1 Transfer Learning.....	45
2.5.2 Μετρήσεις Αξιολόγησης.....	50
2.5.3 Προ-εκπαιδευμένο μοντέλο VGG16.....	51
2.5.4 Προ-εκπαιδευμένο μοντέλο ResNet50	56
2.5.5 Προ-εκπαιδευμένο μοντέλο InceptionV3	60
2.5.6 Model Ensemble	64
2.5.7 Προβλέψεις σε πραγματικό χρόνο	67
Κεφάλαιο 3: Συμπεράσματα και Μελλοντικές Βελτιώσεις.....	68
3.1 Συμπέρασμα.....	68
3.2 Μελλοντικές Αλλαγές και Βελτιώσεις.....	68
Βιβλιογραφία – Αναφορές – Διαδικτυακές Πηγές	70

Κεφάλαιο 1^ο : Εισαγωγή

1.1 Εισαγωγή

Ο σκοπός αυτής της εργασίας είναι να προσφέρει λύση στο πρόβλημα των ατυχημάτων που προκαλούνται από περισπασμένους οδηγούς, μέσω της Βαθιάς Μάθησης. Πιο συγκεκριμένα εκπαιδεύτηκαν τρία νευρωνικά δίκτυα πάνω στο σύνολο δεδομένων «State farm distracted driver detection», στην αναγνώριση της στάσης σώματος του οδηγού και την κατηγοριοποίηση της ανάλογα το είδος του περισπασμού. Έπειτα εξετάζεται και η μέθοδος του ensemble όπου επιτρέπει εκμετάλλευση της δύναμης δύο μοντέλων για την επίτευξη ενός κοινού στόχου.

1.1.1 Παράγοντες Κινδύνου

Μερικοί από τους παράγοντες που οδηγούν σε τροχαία ατυχήματα είναι οι εξής:

- Υπερβολική ταχύτητα.
- Οδήγηση υπό την επήρεια αλκοόλ ή άλλων ψυχοδραστικών ουσιών.
- Απόσπαση της προσοχής του οδηγού.
- Μη ασφαλής οδική υποδομή.
- Απερίσκεπτη οδήγηση.

1.1.2 Απόσπαση της προσοχής του οδηγού

Τί σημαίνει ο όρος «απόσπαση της προσοχής του οδηγού»; Αυτός ο όρος συμπεριλαμβάνει οποιαδήποτε δραστηριότητα μπορεί να αποσπάσει την προσοχή του οδηγού από την ασφαλή οδήγηση. Τέτοιες δραστηριότητες μπορεί να είναι η ομιλία στο κινητό τηλέφωνο ή με κάποιον συνεπιβάτη, ανταλλαγή γραπτών μηνυμάτων μέσω κινητού τηλεφώνου, ο καλλωπισμός, η κατανάλωση αλκοολούχων και μη ποτών, η ενασχόληση με κάποια συσκευή του αυτοκινήτου όπως το ραδιόφωνο και άλλες λοιπές δραστηριότητες, όπως η προσπάθεια του οδηγού να πιάσει κάτι που βρίσκεται στις πίσω θέσεις. [2]

1.1.3 Στατιστικά

Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (World Health Organization), κάθε χρόνο οι ζωές περίπου 1.35 εκατομμυρίων ανθρώπων χάνονται ως αποτέλεσμα τροχαίων ατυχημάτων. Μεταξύ 20 και 50 εκατομμυρίων ανθρώπων υφίστανται μη θανατηφόρους τραυματισμούς, ενώ πολλοί υπέστησαν αναπηρία ως αποτέλεσμα κάποιου τραυματισμού τους [1].

Αφού εξετάστηκαν περισσότερες από 350 μελέτες, η Governors Highway Safety Administration (GHSA), εξήχθη το συμπέρασμα ότι η απόσπαση της προσοχής του οδηγού έχει πιθανότητα 15-25% να είναι η αιτία του ατυχήματος στις ΗΠΑ. Επίσης το NCDOT αναφέρει ότι

στην πολιτεία της Βόρειας Καρολίνας 2.566 ατυχήματα είχαν ως αιτία την απόσπαση της προσοχής του οδηγού το 2010, το οποίο οδήγησαν σε 9 θανάτους και 943 τραυματισμούς. [3]

Σύμφωνα με την National Highway Traffic Safety Administration, στις ΗΠΑ το 2013,

- Το 10% των θανατηφόρων ατυχημάτων, το 18% των ατυχημάτων με τραυματισμούς και το 16% όλων των ατυχημάτων, αναφέρθηκαν από την αστυνομία ως ατυχήματα που προκλήθηκαν από απόσπαση της προσοχής του οδηγού.
- Σκοτώθηκαν 3.154 άτομα και άλλα 424.000 τραυματίστηκαν σε τροχαία ατυχήματα που προκλήθηκαν από απόσπαση της προσοχής του οδηγού.
- Το 10% όλων των οδηγών ηλικίας 15 έως 19 ετών που εμπλέκονται σε θανατηφόρα ατυχήματα αναφέρθηκε ότι είχαν αποσπασμένη την προσοχή τους κατά τη στιγμή των ατυχημάτων. Αυτή η ηλικιακή ομάδα έχει το μεγαλύτερο ποσοστό οδηγών με αποσπασμένη την προσοχή τους τη στιγμή των ατυχημάτων.
- Έχασαν τη ζωή τους 480 μη επιβάτες σε συγκρούσεις που προκλήθηκαν από αποσπασμένη προσοχή. [4]

	Crashes	Drivers	Fatalities
Total	30,057	44,574	32,719
Distraction-Affected (D-A)	2,910 (10% of total crashes)	2,959 (7% of total drivers)	3,154 (10% of total fatalities)

Εικόνα 1: Θανατηφόρα ατυχήματα, Οδηγοί σε θανατηφόρα ατυχήματα και θάνατοι το 2013. [4]

Age Group	Total Drivers		Distraction-Affected Drivers		
	#	% of Total	#	% of Total Drivers	% of Distraction-Affected Drivers
15-19	2,839	6	294	10	10
20-29	10,427	23	803	8	27
30-39	7,598	17	517	7	17
40-49	7,321	16	423	6	14
50-59	7,079	16	384	5	13
60-69	4,483	10	258	6	9
70+	3,951	9	252	6	9
Total	44,574	100	2,959	7	100

Εικόνα 2: Οδηγοί που συμμετέχουν σε θανατηφόρα δυστυχήματα κατά ηλικία και απόσπασης προσοχής. [4]

Σύμφωνα με την ίδια πηγή, στις ΗΠΑ το 2014,

- Το 10% των θανατηφόρων ατυχημάτων, το 18% των τραυματισμών και το 16% όλων των τροχαίων ατυχημάτων, αναφέρθηκαν από την αστυνομία ως ατυχήματα που προκλήθηκαν από οδηγούς με αποσπασμένη την προσοχή τους.
- 3.179 άνθρωποι έχασαν τη ζωή τους και εκτιμάται ότι επιπλέον 431.000 τραυματίστηκαν σε αυτοκινητιστικά ατυχήματα που αφορούσαν οδηγούς με αποσπασμένη την προσοχή.

- Το 10% όλων των οδηγών ηλικίας 15 έως 19 ετών που ενεπλάκησαν σε θανατηφόρα ατυχήματα αναφέρθηκαν ότι είχαν αποσπασμένη την προσοχή τους τη στιγμή των ατυχημάτων. Αυτή η ηλικιακή ομάδα έχει το μεγαλύτερο ποσοστό οδηγών με αποσπασμένη την προσοχή τους τη στιγμή των ατυχημάτων.
- Υπήρξαν 520 μη επιβάτες (πεζοί, ποδηλάτες και άλλοι) που σκοτώθηκαν σε συγκρούσεις που προκλήθηκαν από οδηγούς με αποσπασμένη την προσοχή. [5]

	Crashes	Drivers	Fatalities
Total	29,989	44,583	32,675
Distraction-Affected (D-A)	2,955 (10% of total crashes)	3,000 (7% of total drivers)	3,179 (10% of total fatalities)
Cell Phone in Use	385 (13% of D-A crashes)	398 (13% of distracted drivers)	404 (13% of fatalities in D-A crashes)

Εικόνα 3: Θανατηφόρα ατυχήματα, Οδηγοί σε θανατηφόρα ατυχήματα και θάνατοι το 2014. [5]

Age Group	Total Drivers		Distracted Drivers		
	#	% of Total	#	% of Total Drivers	% of Distracted Drivers
15-19	2,898	7	285	10	10
20-29	10,518	24	859	8	29
30-39	7,573	17	500	7	17
40-49	6,891	15	391	6	13
50-59	7,240	16	404	6	13
60-69	4,560	10	256	6	9
70+	4,027	9	269	7	9
Total	44,583	100	3,000	7	100

Εικόνα 4: Οδηγοί που συμμετέχουν σε θανατηφόρα δυστυχήματα κατά ηλικία και απόσπασης προσοχής. [5]

Σύμφωνα με την ίδια πηγή, στις ΗΠΑ το 2015,

- Το 10% των θανατηφόρων ατυχημάτων, το 15% των τραυματισμών και το 14% όλων των τροχαίων ατυχημάτων, αναφέρθηκαν από την αστυνομία ως ατυχήματα που προκλήθηκαν από οδηγούς με αποσπασμένη την προσοχή τους.
- 3.477 άνθρωποι έχασαν τη ζωή τους και εκτιμάται ότι επιπλέον 391.000 τραυματίστηκαν σε αυτοκινητιστικά ατυχήματα που αφορούσαν οδηγούς με αποσπασμένη την προσοχή.
- Το 9% όλων των οδηγών ηλικίας 15 έως 19 ετών που ενεπλάκησαν σε θανατηφόρα ατυχήματα αναφέρθηκαν ότι είχαν αποσπασμένη την προσοχή τους τη στιγμή των ατυχημάτων. Αυτή η ηλικιακή ομάδα έχει το μεγαλύτερο ποσοστό οδηγών με αποσπασμένη την προσοχή τους τη στιγμή των ατυχημάτων.
- Υπήρξαν 551 μη επιβάτες (πεζοί, ποδηλάτες και άλλοι) που σκοτώθηκαν σε συγκρούσεις που προκλήθηκαν από οδηγούς με αποσπασμένη την προσοχή. [6]

	Crashes	Drivers	Fatalities
Total	32,166	48,613	35,092
Distraction-Affected (D-A)	3,196 (10% of total crashes)	3,263 (7% of total drivers)	3,477 (10% of total fatalities)
Cell Phone in Use	442 (14% of D-A crashes)	456 (14% of distracted drivers)	476 (14% of fatalities in D-A crashes)

Εικόνα 5: Θανατηφόρα ατυχήματα, Οδηγοί σε θανατηφόρα ατυχήματα και θάνατοι το 2015. [6]

Age Group	Total Drivers		Distracted Drivers		
	Number	% of Total Drivers	Number	% of Total Drivers	% of Distracted Drivers
15–19	3,183	7%	290	9%	9%
20–29	11,428	24%	891	8%	27%
30–39	8,479	17%	612	7%	19%
40–49	7,438	15%	482	6%	15%
50–59	7,785	16%	376	5%	12%
60–69	5,012	10%	275	5%	8%
70+	4,255	9%	287	7%	9%
Total	48,613	100%	3,263	7%	100%

Εικόνα 6: Οδηγοί που συμμετέχουν σε θανατηφόρα δυστυχήματα κατά ηλικία και απόσπασης προσοχής. [6]

Σύμφωνα με την ίδια πηγή, στις ΗΠΑ το 2016,

- 9% των θανατηφόρων ατυχημάτων το 2016 αναφέρθηκαν ως ατυχήματα που προκλήθηκαν από οδηγούς με αποσπασμένη την προσοχή τους.
- 3.450 άνθρωποι έχασαν τη ζωή τους σε αυτοκινητιστικά ατυχήματα που αφορούσαν οδηγούς με αποσπασμένη την προσοχή.
- Το 6% όλων των οδηγών που εμπλέκονται σε θανατηφόρα ατυχήματα αναφέρθηκαν ως αποσπασμένη την προσοχή τους κατά τη στιγμή του δυστυχήματος. Το 9% των οδηγών ηλικίας 15 έως 19 ετών που εμπλέκονται σε θανατηφόρα ατυχήματα αναφέρθηκαν ως αποσπασμένοι. Αυτή η ηλικιακή ομάδα έχει το μεγαλύτερο ποσοστό οδηγών που είχαν αποσπασμένη την προσοχή τους τη στιγμή των θανατηφόρων ατυχημάτων.
- Υπήρξαν 562 μη επιβάτες (πεζοί, ποδηλάτες και άλλοι) που σκοτώθηκαν σε συγκρούσεις που προκλήθηκαν από οδηγούς με αποσπασμένη την προσοχή. [7]

	Crashes	Drivers
Total	34,439	51,914
Distraction-Affected (D-A)	3,157 (9% of total crashes)	3,210 (6% of total drivers)
Cell Phone in Use	444 (14% of D-A crashes)	457 (14% of distracted drivers)

Εικόνα 7: Θανατηφόρα ατυχήματα, Οδηγοί σε θανατηφόρα ατυχήματα και θάνατοι το 2016. [7]

Age Group	Total Drivers		Distracted Drivers		
	Number	% of Total Drivers	Number	% of Total Drivers in This Age Group	% of All Distracted Drivers
15–19	3,323	6%	303	9%	9%
20–29	12,335	24%	898	7%	28%
30–39	9,064	17%	586	6%	18%
40–49	7,797	15%	400	5%	12%
50–59	7,945	15%	415	5%	13%
60–69	5,591	11%	288	5%	9%
70+	4,628	9%	282	6%	9%
Total	51,914	100%	3,210	6%	100%

Εικόνα 8: Οδηγοί που συμμετέχουν σε θανατηφόρα δυστυχήματα κατά ηλικία και απόσπασης προσοχής. [7]

Σύμφωνα με την ίδια πηγή, στις ΗΠΑ το 2017,

- Το 9% των θανατηφόρων ατυχημάτων αναφέρθηκαν ως ατυχήματα που προκλήθηκαν από οδηγούς με αποσπασμένη την προσοχή τους.
- 3.166 άνθρωποι έχασαν τη ζωή τους σε αυτοκινητιστικά ατυχήματα που αφορούσαν οδηγούς με αποσπασμένη την προσοχή.
- Το 6% όλων των οδηγών που ενεπλάκησαν σε θανατηφόρα ατυχήματα αναφέρθηκαν ότι είχαν αποσπασμένη την προσοχή τους τη στιγμή των ατυχημάτων. Το 8% των οδηγών ηλικίας 15 έως 19 ετών που συμμετείχαν σε θανατηφόρα ατυχήματα αναφέρθηκαν ως αποσπασμένοι. Αυτή η ηλικιακή ομάδα έχει το μεγαλύτερο ποσοστό οδηγών που είχαν αποσπασμένη την προσοχή τους τη στιγμή των θανατηφόρων ατυχημάτων.
- Υπήρξαν 599 μη επιβάτες (πεζοί, ποδηλάτες και άλλοι) που σκοτώθηκαν σε συγκρούσεις που προκλήθηκαν από οδηγούς με αποσπασμένη την προσοχή. [8]

	Crashes	Drivers	Fatalities
Total	34,247	52,274	37,133
Distraction-Affected (D-A)	2,935 (9% of total crashes)	2,994 (6% of total drivers)	3,166 (9% of total fatalities)
Cell Phone in Use	401 (14% of D-A crashes)	404 (13% of distracted drivers)	434 (14% of fatalities in D-A crashes)

Εικόνα 9: Θανατηφόρα ατυχήματα, Οδηγοί σε θανατηφόρα ατυχήματα και θάνατοι το 2017. [8]

Age Group	Total Drivers		Distracted Drivers		
	Number	% of Total Drivers	Number	% of Total Drivers in This Age Group	% of All Distracted Drivers
15–19	3,255	6%	271	8%	9%
20–29	12,086	23%	816	7%	27%
30–39	9,290	18%	557	6%	19%
40–49	7,944	15%	431	5%	14%
50–59	8,029	15%	360	4%	12%
60–69	5,562	11%	224	4%	7%
70+	4,911	9%	292	6%	10%
Total	52,274	100%	2,994	6%	100%

Εικόνα 10: Οδηγοί που συμμετέχουν σε θανατηφόρα δυστυχήματα κατά ηλικία και απόσπασης προσοχής. [8]

Σύμφωνα με την ίδια πηγή, στις ΗΠΑ το 2018,

- Το 8% των θανατηφόρων ατυχημάτων, 15% των ατυχημάτων με τραυματισμούς και το 14% όλων των ατυχημάτων, αναφέρθηκαν από την αστυνομία ως ατυχήματα που προκλήθηκαν από οδηγούς με αποσπασμένη προσοχή.
- 400.000 άτομα τραυματίστηκαν ενώ άλλα 2.841 άφησαν την τελευταία του πνοή σε τροχαία ατυχήματα που προκλήθηκαν από οδηγούς που τους είχε αποσπαστεί η προσοχή.
- Το 5% όλων των οδηγών που εμπλέκονται σε θανατηφόρα ατυχήματα αναφέρθηκαν ως αποσπασμένοι κατά τη στιγμή των ατυχημάτων. Το 8% των οδηγών ηλικίας 15 έως 19 ετών που συμμετείχαν σε θανατηφόρα δυστυχήματα αναφέρθηκαν ως αποσπασμένοι. Αυτή η ηλικιακή ομάδα έχει τη μεγαλύτερη αναλογία οδηγών που είχαν αποσπασμένη την προσοχή τους κατά τη διάρκεια των θανατηφόρων ατυχημάτων.
- Υπήρξαν 506 μη επιβάτες (πεζοί, ποδηλάτες και άλλοι) που σκοτώθηκαν σε συγκρούσεις που επηρεάστηκαν από οδηγούς με αποσπασμένη την προσοχή τους.

	Total	Distraction-Affected (D-A)	
		Number	Percentage of Total
Crashes	33,654	2,628	8%
Drivers	51,490	2,688	5%
Fatalities	36,560	2,841	8%

Εικόνα 11: Θανατηφόρα ατυχήματα, Οδηγοί σε θανατηφόρα ατυχήματα και θάνατοι το 2018. [9]

Age Group	Total Drivers		Distracted Drivers		
	Number	Percentage of Total Drivers	Number	Percentage of Total Drivers in This Age Group	Percentage of All Distracted Drivers
15-19	3,070	6%	237	8%	9%
20-29	11,629	23%	680	6%	25%
30-39	9,264	18%	506	5%	19%
40-49	7,809	15%	389	5%	14%
50-59	7,883	15%	342	4%	13%
60-69	5,635	11%	251	4%	9%
70+	4,919	10%	256	5%	10%
Total	51,490	100%	2,688	5%	100%

Εικόνα 12: Οδηγοί που συμμετέχουν σε θανατηφόρα δυστυχήματα κατά ηλικία και απόσπασης προσοχής. [9]

Σύμφωνα με την ίδια πηγή, στις ΗΠΑ το 2019,

- Το 9% των θανατηφόρων ατυχημάτων, το 15% των τραυματισμών και το 15% όλων των τροχαίων ατυχημάτων, αναφέρθηκαν από την αστυνομία ως ατυχήματα που προκλήθηκαν από οδηγούς με αποσπασμένη προσοχή.
- 3.142 άνθρωποι έχασαν τη ζωή τους και εκτιμάται ότι επιπλέον 424.000 άνθρωποι τραυματίστηκαν σε τροχαία ατυχήματα που αφορούσαν οδηγούς με αποσπασμένη την προσοχή τους.
- Το 6% όλων των οδηγών που εμπλέκονται σε θανατηφόρα ατυχήματα, αναφέρθηκε ότι είχαν αποσπασμένη την προσοχή τους τη στιγμή των ατυχημάτων. Το 9% των οδηγών ηλικίας 15 έως 20 ετών που συμμετείχαν σε θανατηφόρα ατυχήματα αναφέρθηκαν ως αποσπασμένοι. Αυτή η ηλικιακή ομάδα έχει το μεγαλύτερο ποσοστό οδηγών που είχαν αποσπασμένη την προσοχή τους τη στιγμή των θανατηφόρων ατυχημάτων.
- Υπήρξαν 566 μη επιβάτες (πεζοί, ποδηλάτες και άλλοι) που σκοτώθηκαν σε συγκρούσεις που επηρεάστηκαν από οδηγούς με αποσπασμένη την προσοχή τους. [10]

	Total	Distraction-Affected (D-A)	
		Number	Percentage of Total
Crashes	33,244	2,895	9%
Drivers Involved	50,930	3,008	6%
Fatalities	36,096	3,142	9%

Εικόνα 13: Θανατηφόρα ατυχήματα, Οδηγοί σε θανατηφόρα ατυχήματα και θάνατοι το 2019. [10]

Age Group	Total Drivers		All Distracted Drivers		
	Number	Percentage of Total Drivers	Number	Percentage of Total Drivers in This Age Group	Percentage of All Distracted Drivers
15-20	3,968	8%	344	9%	11%
21-24	4,590	9%	297	6%	10%
25-34	10,507	21%	686	7%	23%
35-44	8,301	16%	531	6%	18%
45-54	7,532	15%	392	5%	13%
55-64	7,166	14%	315	4%	10%
65-74	4,404	9%	208	5%	7%
75+	3,229	6%	184	6%	6%
Total	50,930	100%	3,008	6%	100%

Εικόνα 14: Οδηγοί που συμμετέχουν σε θανατηφόρα δυστυχήματα κατά ηλικία και απόσπασης προσοχής. [10]

1.1.4 Συμπέρασμα

Από τα παραπάνω στατιστικά συμπεραίνουμε ότι το πρόβλημα είναι αρκετά σοβαρό και χρήζει αντιμετώπισης. Ο τρόπος που προτείνεται σε αυτή τη διπλωματική εργασία δεν είναι άλλος από την τεχνίτη νοημοσύνη. Ήτοι με τη χρήση νευρωνικών δικτύων που θα εκπαιδευτούν ώστε να αναγνωρίζουν πότε ο οδηγός έχει στρέψει την προσοχή του σε άλλη δραστηριότητα, πέρα της ασφαλούς οδήγησης, καθώς και το είδος της. Βέβαια για να γίνει κάτι τέτοιο χρειαζόμαστε δεδομένα, για τα οποία θα γίνουν αναφορές παρακάτω.

1.1.5 Άλλα συστήματα που αυξάνουν την ασφάλεια στην οδήγηση

Έκτος από την λύση της τεχνίτης νοημοσύνης για την ταξινόμηση της συμπεριφοράς του οδηγού, υπάρχουν κι' άλλα συστήματα που αυξάνουν την ασφάλεια στην οδήγηση. Τέτοια συστήματα ανήκουν στην κατηγορία των «προηγμένων συστημάτων υποβοήθησης οδηγού» (Advanced driver-assistance systems - ADAS) και βοηθούν τους οδηγούς σε λειτουργίες οδήγησης και στάθμευσης. Τέτοια συστήματα επικοινωνούν με τον χρήστη μέσω μια διεπαφής ανθρώπου-μηχανής, ενώ λαμβάνουν ερεθίσματα από το περιβάλλον χρησιμοποιώντας αυτοματοποιημένη τεχνολογία όπως αισθητήρες και κάμερες έτσι ώστε να ανταποκρίνονται ανάλογα.

Μερικά από αυτά τα συστήματα:

- Adaptive cruise control (ACC). Μπορεί να διατηρήσει μια επιλεγμένη ταχύτητα και απόσταση μεταξύ ενός οχήματος και του προπορευόμενου οχήματος. Μπορούν επίσης να φρενάρουν, να επιταχύνουν ακόμα και να σταματήσουν αν χρειαστεί. Ωστόσο απαιτούν την προσοχή του οδηγού την ώρα που είναι ενεργό.
- Συσκευές κλειδώματος ανάφλεξης για αλκοόλ. Δεν αφήνουν τους οδηγούς να ξεκινήσουν το αυτοκίνητο εάν το επίπεδο αλκοόλ στην αναπνοή τους είναι πάνω από το επιτρεπτό όριο.
- Οθόνη head-up αυτοκινήτου (auto-HUD). Εμφανίζει στον οδηγό τις βασικές πληροφορίες του συστήματος σε πλεονεκτικό σημείο που δεν απαιτεί από τον οδηγό να κοιτάζει κάτω ή μακριά από το δρόμο.
- Οθόνη τυφλού σημείου. Περιλαμβάνει κάμερες που παρακολουθούν τα τυφλά σημεία του οδηγού και ειδοποιούν τον οδηγό εάν τυχόν εμπόδια πλησιάσουν το όχημα. Τα τυφλά σημεία ορίζονται ως οι περιοχές πίσω ή στο πλάι του οχήματος που ο οδηγός δεν μπορεί να δει από τη θέση του.
- Το σύστημα αποφυγής σύγκρουσης. Χρησιμοποιεί μικρούς ανιχνευτές ραντάρ, συνήθως τοποθετημένους στο μπροστινό μέρος του αυτοκινήτου, για να καθορίσει την εγγύτητα του αυτοκινήτου με κοντινά εμπόδια και να ειδοποιήσει τον οδηγό για πιθανή σύγκρουση.
- Ανίχνευση υπνηλίας οδηγού. Το σύστημα λαμβάνει διάφορες πληροφορίες όπως μοτίβα προσώπου, συνήθειες οδήγησης και άλλα, προκειμένου να διαπιστώσει αν οι δραστηριότητές του αντιστοιχούν στην υπνηλία. Σε αυτή την περίπτωση το σύστημα θα προσπαθήσει να ειδοποιήσει τον οδηγό μέσω δυνατού ήχου ή δόνησης.
- Προειδοποίηση αναχώρησης λωρίδας κυκλοφορίας. Βοηθά τον οδηγό να κρατήσει το αυτοκίνητο στη μέση της λωρίδας. Ένα τέτοιο σύστημα διαβάζει τις σημάνσεις των λωρίδων και αν το όχημα αρχίζει να βγαίνει εκτός πορείας θα ειδοποιήσει τον οδηγό. Κάποια συστήματα μπορούν να αναλάβουν αυτόματα και το τιμόνι.
- Βοήθεια για αλλαγή λωρίδας. Βοηθά τον οδηγό να πραγματοποιήσει αλλαγή λωρίδας με ασφάλεια χρησιμοποιώντας αισθητήρες για να σαρώσει το περιβάλλον του αυτοκινήτου καθώς και να παρακολουθήσει τα τυφλά σημεία του οδηγού. Σε περίπτωση που εντοπίσει κάτι την ώρα της αλλαγής λωρίδας θα ειδοποιήσει τον οδηγό με κάποιο οπτικό ή ακουστικό ερέθισμα.

Αυτά είναι μόνο μερικά από τα συστήματα που χρησιμοποιούνται στα σύγχρονα αυτοκίνητα. [11]

1.2 Τα Δεδομένα

Στη Μηχανική Μάθηση (Machine Learning) και πιο συγκεκριμένα στην Βαθιά Μάθηση (Deep Learning), χρειάζονται δεδομένα για να εκπαιδευτούν τα μοντέλα ή τα νευρωνικά δίκτυα. Το σύνολο δεδομένων που χρησιμοποιήθηκε είναι το «State Farm Distracted Driver Detection».

1.2.1 Το σύνολο δεδομένων “State Farm Distracted Driver Detection”.

Πρόκειται για ένα πακέτο δεδομένων (dataset) που είχε δημοσιευτεί από τον όμιλο ασφαλιστικών εταιριών State Farm Insurance, ως μέρος ενός διαγωνισμού με έπαθλο 65.000\$, στην ιστοσελίδα του Kaggle. Το Kaggle είναι μια διαδικτυακή κοινότητα επιστημόνων και μη, που ασχολούνται με τη μηχανική μάθηση. Σε αυτή την ιστοσελίδα μπορεί να βρει κανείς, ποικιλία από δεδομένα, όπως αυτό που προαναφέρθηκε, καθώς και διαδικτυακά μαθήματα.

1.2.2 Περιγραφή των δεδομένων

Το σύνολο δεδομένων περιέχει εικόνες από οδηγούς, που έχουν τραβηχτεί από το εσωτερικό ενός αυτοκινήτου και το αποτυπώνουν να εκτελεί κάποια δραστηριότητα. Ο σκοπός είναι να προβλεφθεί το είδος της δραστηριότητας αυτής. [12]

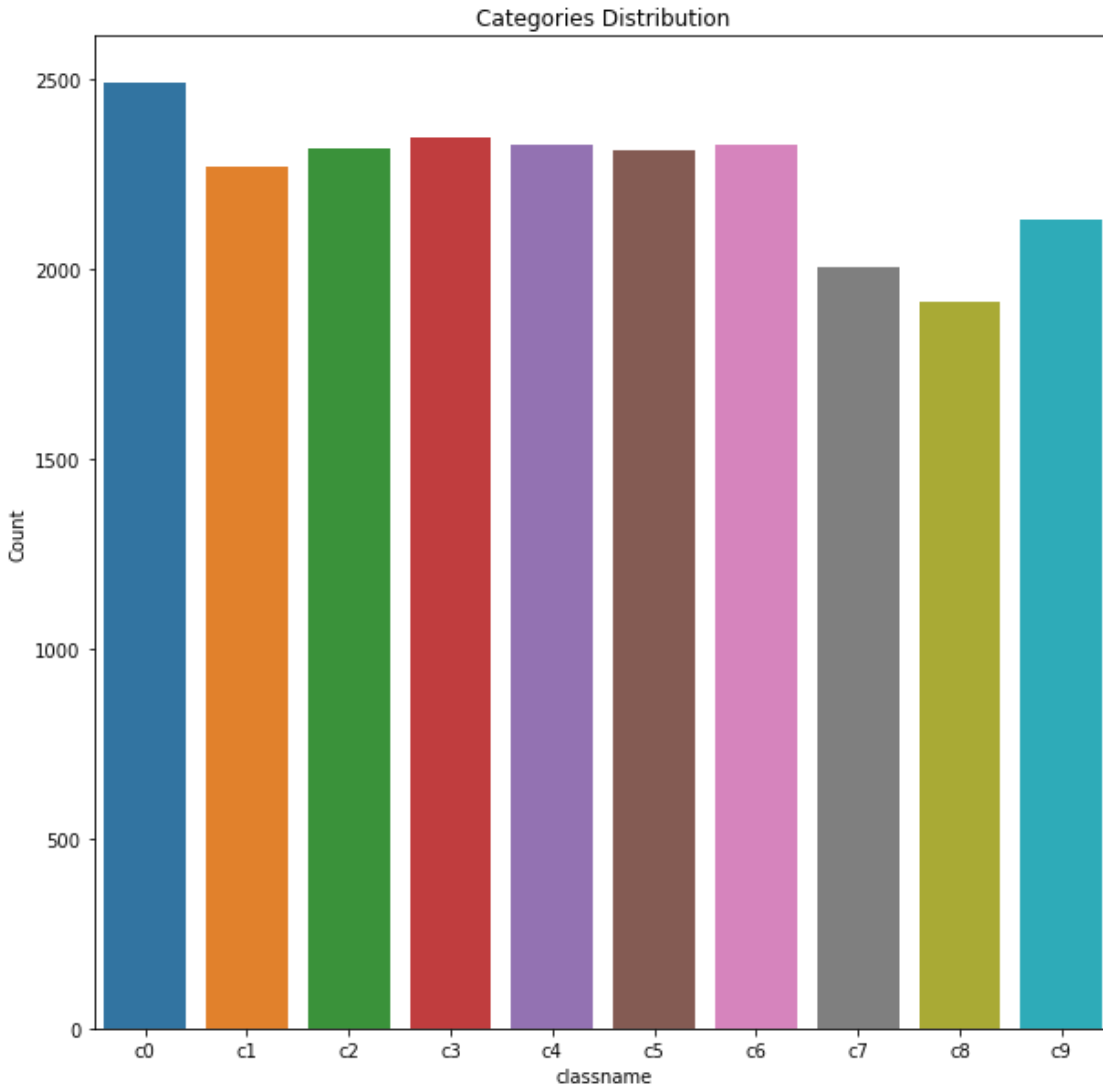
Υπάρχουν δέκα (10) διαφορετικές δραστηριότητες - κατηγορίες, οι οποίες είναι οι εξής:

- c0: safe driving
- c1: texting – right
- c2: talking on the phone – right
- c3: texting – left
- c4: talking on the phone – left
- c5: operating the radio
- c6: drinking
- c7: reaching behind
- c8: hair and makeup
- c9: talking to passenger

1.2.3 Στατιστικά

Το πακέτο δεδομένων περιέχει 22.424 φωτογραφίες χωρισμένες σε φακέλους. Κάθε φάκελος αντιπροσωπεύει μια δραστηριότητα - κατηγορία.

Η κατανομή των φωτογραφιών στις κατηγορίες που ανήκουν φαίνεται στο παρακάτω διάγραμμα.



Εικόνα 15: Διανομή Εικόνων ανά Κατηγορία

Αυτό που παρατηρούμε αμέσως είναι ότι υπάρχει μια ανισορροπία στη διανομή των εικόνων ανά κατηγορία. Αυτό σημαίνει ότι κάποιες κατηγορίες έχουν περισσότερες εικόνες από κάποιες άλλες, κάτι που μπορεί να προκαλέσει προβλήματα. Ήτοι, όσο λιγότερα δεδομένα έχουμε για μια κατηγορία, τόσο μεγαλύτερη είναι η πιθανότητα το μοντέλο μας να «μπερδέψει» αυτή τη κατηγορία με μία άλλη που έχει περισσότερα δεδομένα – εικόνες. Η αντιμετώπιση του προβλήματος αυτού εξετάζεται στο Κεφάλαιο 2.

Ύστερα υπολογίστηκε το πόσες φορές εμφανίζεται ο κάθε οδηγός στις εικόνες του dataset.

	driver_id	Counts
0	p021	1237
1	p022	1233
2	p024	1226
3	p026	1196
4	p016	1078
5	p066	1034
6	p049	1011
7	p051	920
8	p014	876
9	p015	875
10	p035	848
11	p047	835
12	p012	823
13	p081	823
14	p064	820
15	p075	814
16	p061	809
17	p056	794
18	p050	790
19	p052	740
20	p002	725
21	p045	724
22	p039	651
23	p041	605
24	p042	591
25	p072	346

Εικόνα 16: Η Συχνότητα Εμφάνισης των Οδηγών ανά Εικόνα

Στον παραπάνω πίνακα φαίνεται το πόσο συχνά εμφανίζεται ο κάθε οδηγός στις εικόνες του dataset. Όταν κάποιος οδηγός εμφανίζεται πολύ περισσότερο από κάποιους άλλους, το μοντέλο μας μπορεί να «εκλάβει» το πρόσωπό του ως χαρακτηριστικό (feature), κάτι που προφανώς είναι λάθος. Το μοναδικό χαρακτηριστικό που πρέπει να υπολογίζει το μοντέλο είναι η στάση του σώματος του οδηγού.

Στην παρακάτω εικόνα απεικονίζεται ένα παράδειγμα για την κάθε κατηγορία.



Εικόνα 17: Επισκόπηση Εικόνων

1.3 Τεχνητή Νοημοσύνη (Artificial Intelligence)

1.3.1 Σύντομη Ιστορική Αναδρομή

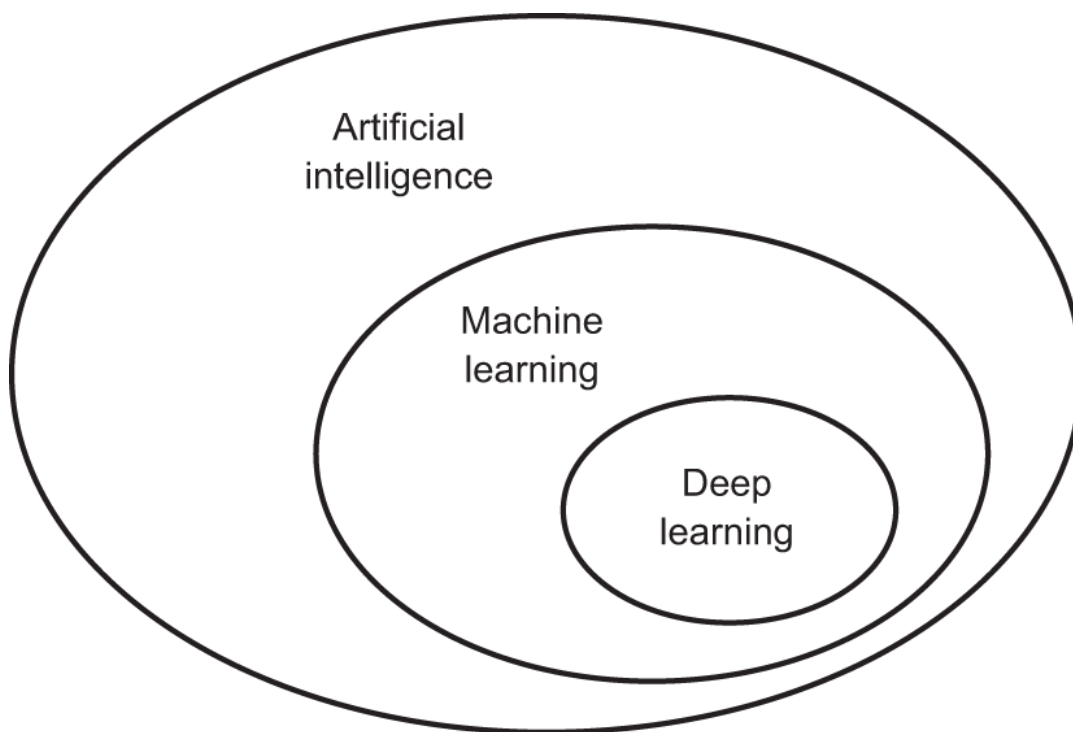
Η Τεχνητή Νοημοσύνη (AI) γεννήθηκε στα τη δεκαετία του 1950, όταν κάποιοι πρωτοπόροι στο νεοεμφανιζόμενο πεδίο της επιστήμης των υπολογιστών άρχισε να αναρωτιέται αν μπορούν οι υπολογιστές να «σκεφτούν». Μια ερώτηση της οποίας οι συνέπειες αποτελούν θέμα για έρευνα μέχρι και σήμερα. [13]

1.3.2 Τί είναι η Τεχνητή Νοημοσύνη (AI);

Η Τεχνητή Νοημοσύνη (AI) μπορεί να περιγράψει ως η προσπάθεια αυτοματοποίησης πνευματικών εργασιών που συνήθως εκτελούνται από άνθρωπο.

Η Τεχνητή Νοημοσύνη (AI) είναι ένα γενικό πεδίο που περιλαμβάνει τα «υποσύνολα» της Μηχανικής Μάθησης και της Βαθιάς Μάθησης. Μέχρι τα τέλη της δεκαετία του 1980, η προσεγγίσεις που υπήρχαν για την Τεχνητή Νοημοσύνη δεν περιλάμβαναν την μάθηση. Δηλαδή τα πρώτα προγράμματα που δημιουργήθηκαν περιείχαν κώδικα με κανόνες για τον χειρισμό γνώσεων που ήταν αποθηκευμένες σε βάσει δεδομένων. Αυτή η προσέγγιση ονομάζεται symbolic AI (συμβολική Τεχνητή Νοημοσύνη).

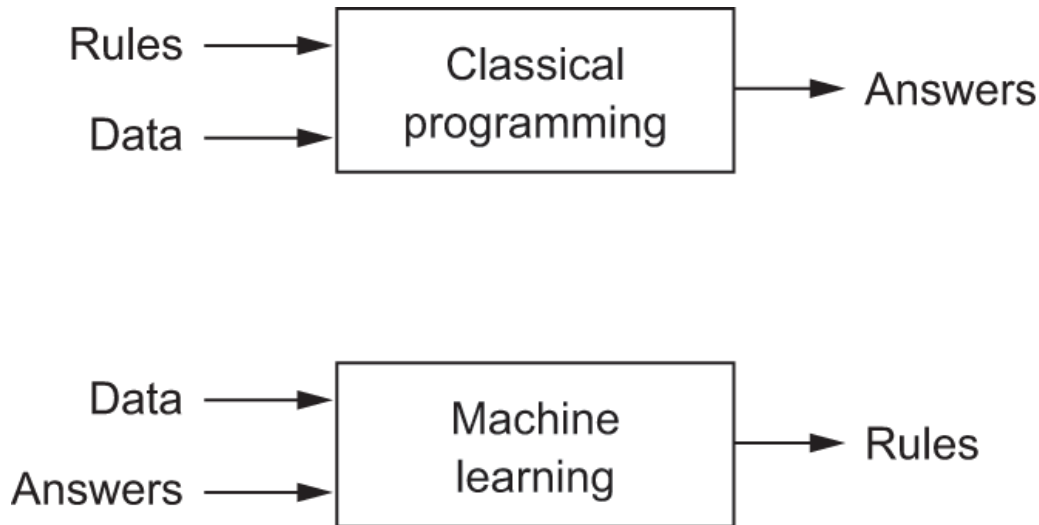
Αργότερα όμως παρουσιάστηκαν ασαφή προβλήματα που η συμβολική Τεχνητή Νοημοσύνη δεν μπορούσε να λύσει και έτσι μια νέα προσέγγιση εμφανίστηκε, αυτή της Μηχανικής Μάθησης. [13]



Εικόνα 18: Τεχνητή νοημοσύνη, μηχανική μάθηση και βαθιά μάθηση. [13]

1.3.3 Μηχανική Μάθηση

Ένα σύστημα μηχανικής μάθησης εκπαιδεύεται πάνω σε δεδομένα-παραδείγματα που του τροφοδοτούν οι άνθρωποι και δεν προγραμματίζεται. Τα δεδομένα αυτά σχετίζονται με την εργασία που πρέπει μάθει να εκτελεί ο υπολογιστής. Το σύστημα «μαθαίνει» βρίσκοντας την στατιστική δομή των δεδομένων. Κατ' αυτόν τον τρόπο, βρίσκει μόνο του τους κανόνες που θα οδηγήσουν στην αυτοματοποίηση της εργασίας. [13] Αντίθετα στον κλασσικό προγραμματισμό, οι κανόνες που θα οδηγήσουν στην αυτοματοποίηση της εργασίας, καθορίζονται σύμφωνα με τη λογική και τη κατανόηση του προβλήματος που έχει ο προγραμματιστής.



Εικόνα 19: Μηχανική Μάθηση & Κλασσικός Προγραμματισμός [13]

Ένα παράδειγμα προβλήματος Μηχανικής Μάθησης είναι η αυτοματοποίηση της προσθήκης λεζάντας σε φωτογραφίες σκύλων ανάλογα την ράτσα. Στο σύστημα θα δοθούν παραδείγματα με εικόνες μαζί με τις ετικέτες τους, που έχουν τοποθετηθεί από άνθρωπο και το αυτό θα μάθει τους στατιστικούς κανόνες για την σύνδεση συγκεκριμένων εικόνων με συγκεκριμένες ετικέτες.

1.3.3.1 Εκμάθηση κανόνων και αναπαραστάσεων από δεδομένα

Για την πλήρη κατανόηση του τρόπου λειτουργίας των αλγορίθμων μηχανικής μάθησης, κρίνεται χρήσιμο να γίνει αναφορά στην μορφή των δεδομένων που τροφοδοτούνται. Όπως αναφέρθηκε και παραπάνω οι εν λόγω αλγόριθμοι χρειάζονται δεδομένα για να βρουν τους κανόνες που θα οδηγήσουν στην αυτοματοποίηση. Τα δεδομένα αυτά έχουν τα εξής χαρακτηριστικά:

- Δεδομένα εισόδου. Για παράδειγμα, σε προβλήματα μηχανικής όρασης, τα δεδομένα εισόδου είναι εικόνες.
- Παραδείγματα της αναμενόμενης εξόδου. Για παράδειγμα, σε προβλήματα μηχανικής όρασης, είναι ετικέτες που αντιστοιχούν σε κάθε εικόνα όπως «σκύλος», «γάτα», «άνθρωπος».

- Τρόπος μέτρησης των επιδόσεων του αλγορίθμου. Είναι απαραίτητο να μετρηθεί η απόσταση της τρέχουσας εξόδου του αλγορίθμου από την αναμενόμενη, διότι θα χρησιμοποιηθεί ως σήμα ανατροφοδότησης για την βελτίωση της λειτουργίας του.

Εν κατακλείδι, η δουλειά που έχει να κάνει ένας τέτοιος αλγόριθμος είναι να μετατρέψει τα δεδομένα εισόδου σε αναπαραστάσεις σχετικές με την αναμενόμενη έξοδο. Αυτό το μαθαίνει εκθέτοντάς το σε παραδείγματα δεδομένων εισόδου με τις αντίστοιχες εξόδους τους. [13]

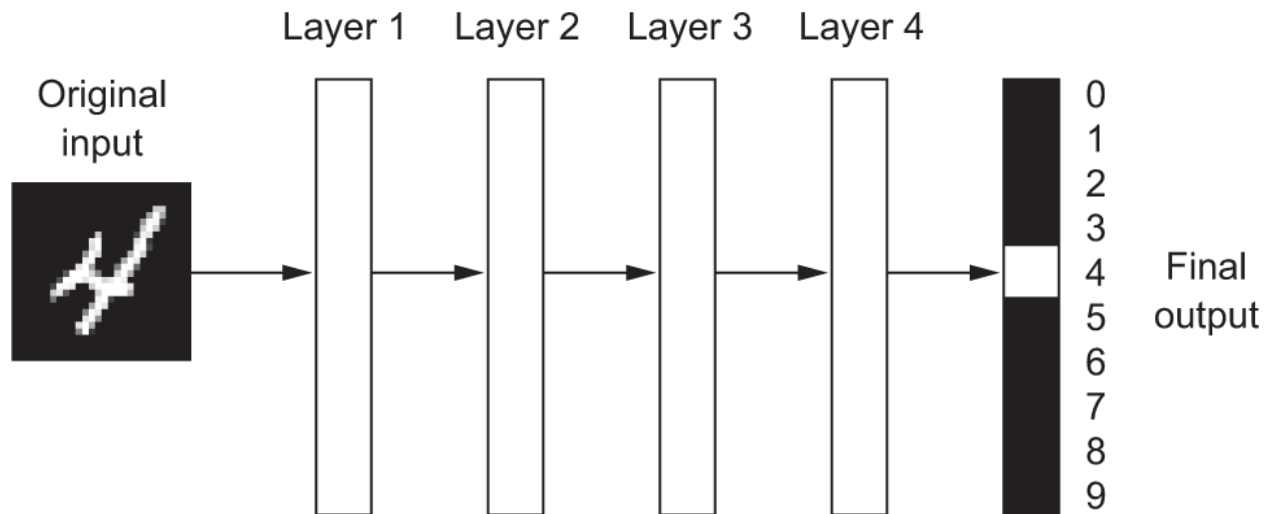
Τί είναι όμως μια αναπαράσταση; Αναπαράσταση είναι ένας τρόπος αναπαράστασης-κωδικοποίησης των δεδομένων. Οι αλγόριθμοι μηχανικής μάθησης μετατρέπουν τα δεδομένα εισόδου σε μια μορφή που θα είναι πιο διακριτά τα χρήσιμα χαρακτηριστικά τους για την επίτευξη του στόχου που έχει τεθεί.

1.3.3 Βαθιά Μάθηση

Η Βαθιά Μάθηση είναι ένα υποπεδίο της Μηχανικής Μάθησης και αποτελεί μια διαφορετική εκδοχή των μαθησιακών αναπαραστάσεων των δεδομένων, που δίνει έμφαση στη μάθηση διαδοχικών επιπέδων ολοένα και πιο σημαντικών αναπαραστάσεων. Με την λέξη «βαθιά» δεν εννοείται η κατανόηση που επιτυγχάνεται από την προσέγγιση, αλλά από τα πολυάριθμα διαδοχικά επίπεδα αναπαραστάσεων των δεδομένων. Ο αριθμός των επιπέδων αυτών, ονομάζεται «βάθος του μοντέλου».

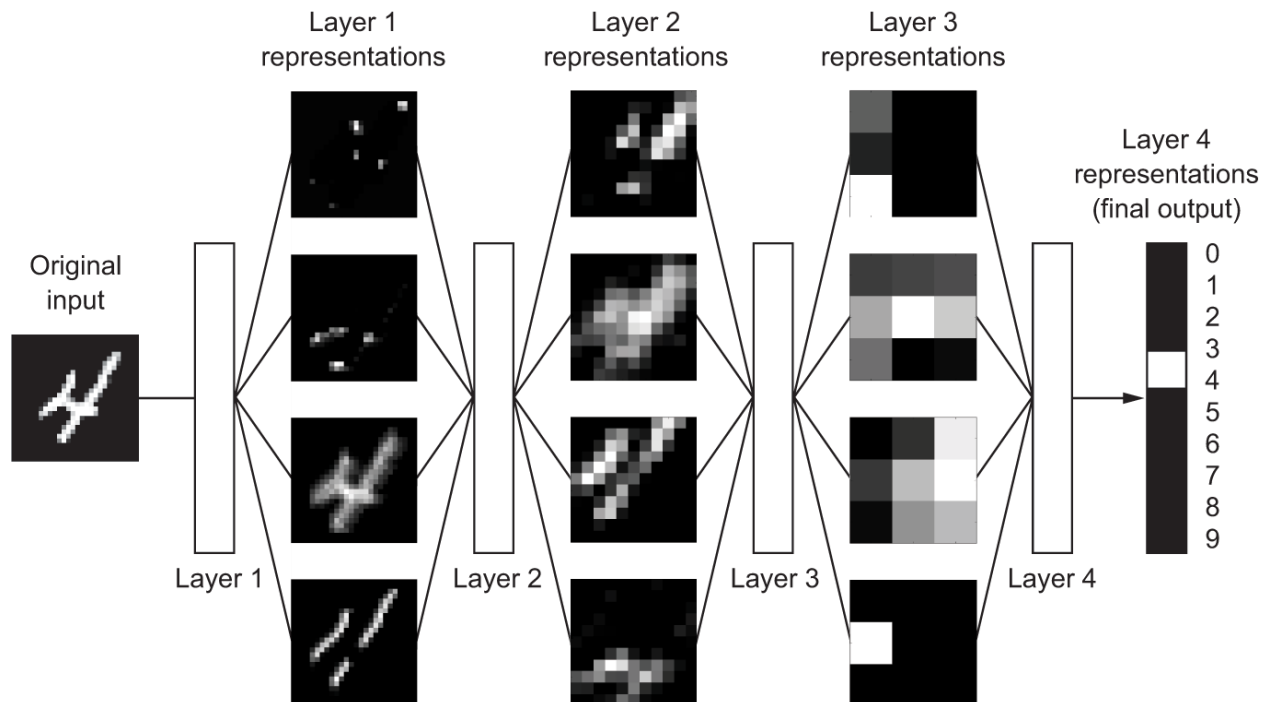
Στην Βαθιά Μάθηση τα μοντέλα αυτά λέγονται «νευρωνικά δίκτυα», δομημένα σε κυριολεκτικά στρώματα τα οποία στοιβάζονται στο ένα πάνω στο άλλο. Ο όρος νευρωνικό δίκτυο είναι μια αναφορά στη Νευροβιολογία, και αυτό γιατί κάποιες έννοιες της Βαθιάς Μάθησης είναι εμπνευσμένες από τον ανθρώπινο εγκέφαλο. Σε καμία όμως περίπτωση δεν υποστηρίζεται ότι ο εγκέφαλος εφαρμόζει τεχνικές της Βαθιάς Μάθησης για να μάθει.

Πώς μοιάζουν οι αναπαραστάσεις που μαθαίνει ένας αλγόριθμος Βαθιάς Μάθησης; Ακολουθεί ένα παράδειγμα (Εικόνα 20) με ένα δίκτυο πολλών επιπέδων το οποίο μετατρέπει μια εικόνα ενός ψηφίου για να το αναγνωρίσει.



Εικόνα 20: Ένα Βαθύ Νευρωνικό Δίκτυο για αναγνώριση ψηφίων [13]

Όπως μπορείτε να δείτε στην Εικόνα 21, το δίκτυο μετατρέπει την ψηφιακή εικόνα σε αναπαραστάσεις που διαφέρουν όλο και περισσότερο από την αρχική εικόνα, ωστόσο είναι όλο και πιο ενημερωτικές για το τελικό αποτέλεσμα. Ένα βαθύ δίκτυο μπορεί να παρομοιαστεί με μια λειτουργία απόσταξης πληροφοριών πολλαπλών σταδίων, όπου οι πληροφορίες περνούν από διαδοχικά φίλτρα και βγαίνουν όλο και περισσότερο απλές.



Εικόνα 21: Αναπαραστάσεις που έχει μάθει ένα μοντέλο κατηγοριοποίησης ψηφίων [13]

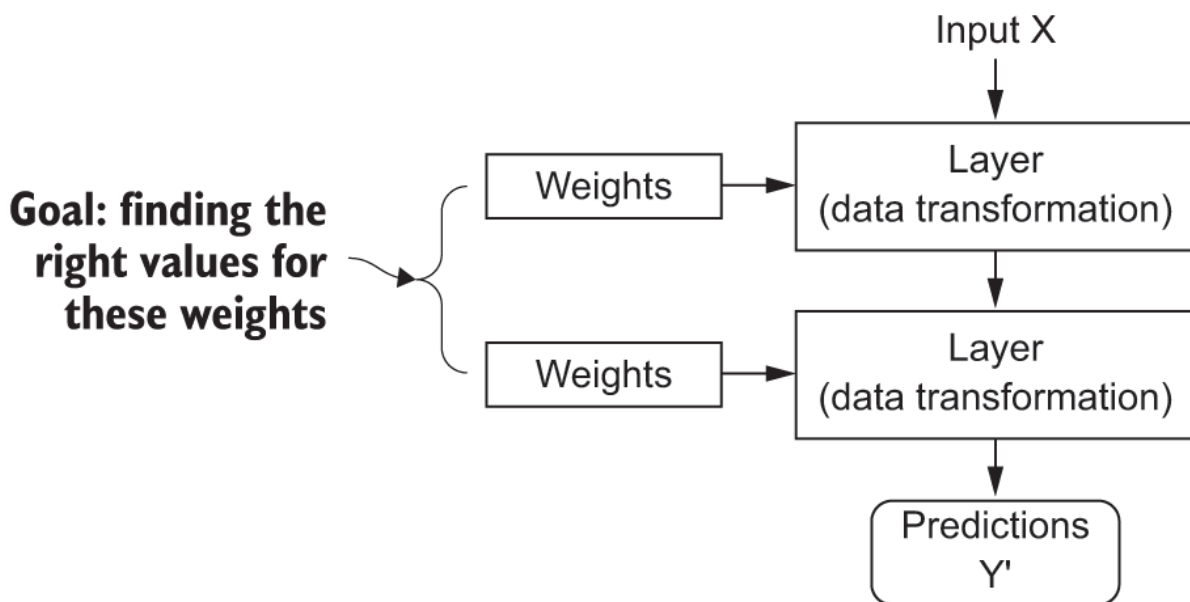
Εν κατακλείδι η βαθιά μάθηση είναι ένας πολυβάθμιος τρόπος για να μάθει ένα μοντέλο αναπαραστάσεις δεδομένων.

1.3.4 Πώς λειτουργεί ένα νευρωνικό δίκτυο;

Η βασική δομή δεδομένων στα νευρωνικά δίκτυα είναι το στρώμα ή επίπεδο. Ένα στρώμα είναι ένα δομοστοιχείο επεξεργασίας δεδομένων που λαμβάνει ως είσοδο και έχει ως έξοδο έναν ή περισσότερους ταυστές. Μερικά στρώματα μπορεί να μην έχουν κατάσταση, αλλά πιο συχνά στρώματα έχουν κατάσταση: τα βάρη του επιπέδου, μαθαίνουν έναν ή περισσότερους ταυστές με στοχαστική κάθοδο κλίσης, τα οποία όλα μαζί περιέχουν τις γνώσεις του δικτύου.

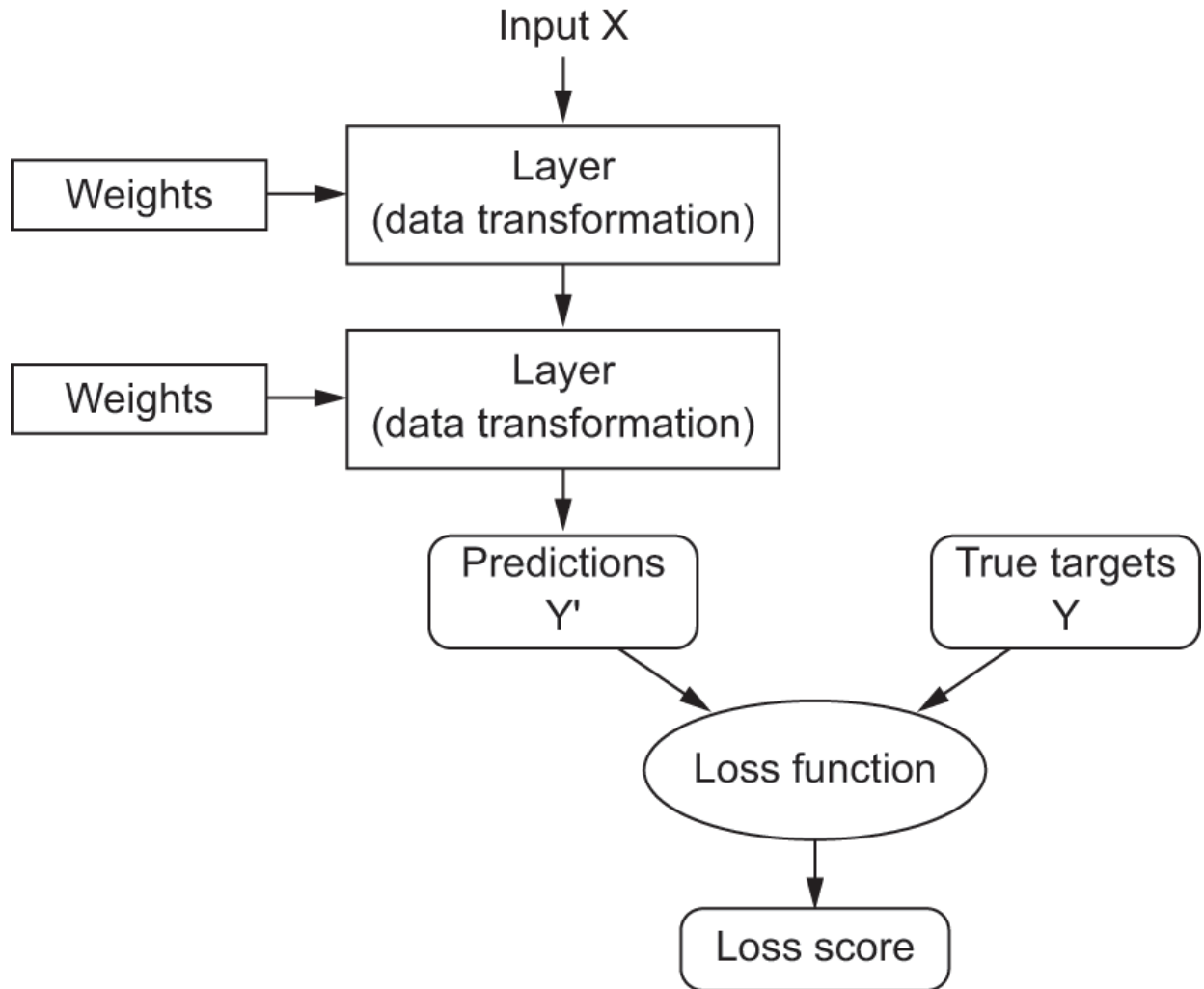
Διαφορετικοί τύποι στρωμάτων είναι κατάλληλα για διαφορετικές μορφές ταυστών και τα διάφορα είδη επεξεργασίας δεδομένων. Για παράδειγμα, απλά διανυσματικά δεδομένα αποθηκευμένα σε δυσδιάστατους ταυστές, συχνά υποβάλλονται σε επεξεργασία από πυκνά συνδεδεμένα στρώματα, που ονομάζονται επίσης πλήρως συνδεδεμένα ή πυκνά στρώματα (Dense Layers). Αντίστοιχα, δεδομένα αλληλουχίας που αποθηκεύονται σε τρισδιάστατους ταυστές, τυπικά υποβάλλονται σε επεξεργασία από επαναλαμβανόμενα επίπεδα (recurrent layers), όπως ένα στρώμα LSTM (Long Short-Term Memory) ή μονοδιάστατα στρώματα συνέλιξης (1D convolution layers). Τέλος, τα δεδομένα εικόνας που αποθηκεύονται σε τετραδιάστατους ταυστές, υποβάλλονται συνήθως σε επεξεργασία με στρώματα δυσδιάστατης συνέλιξης (2D convolution layers). [13]

Οι αλλαγές που υπόκεινται οι είσοδοι κάθε επιπέδου εξαρτάται από έναν αριθμό που λέγεται βάρος – weight. Οι τιμές των βαρών δεν μένουν σταθερές, αλλά κατά τη διάρκεια της εκμάθησης οι τιμές αυτές αλλάζουν συνεχώς στην προσπάθεια εύρεσης του καταλληλότερου συνόλου τιμών για τα βάρη, ώστε να έρθει πιο κοντά στον στόχο του. Ένα νευρωνικό δίκτυο μπορεί να περιέχει δεκάδες εκατομμύρια παραμέτρους, γι' αυτό το λόγο άλλωστε λέγονται βαθιά. Η εύρεση αυτού του συνόλου τιμών αποτελεί μια δύσκολη διαδικασία, μιας και κάθε αλλαγή επηρεάζει όλες τις υπόλοιπες. [13]



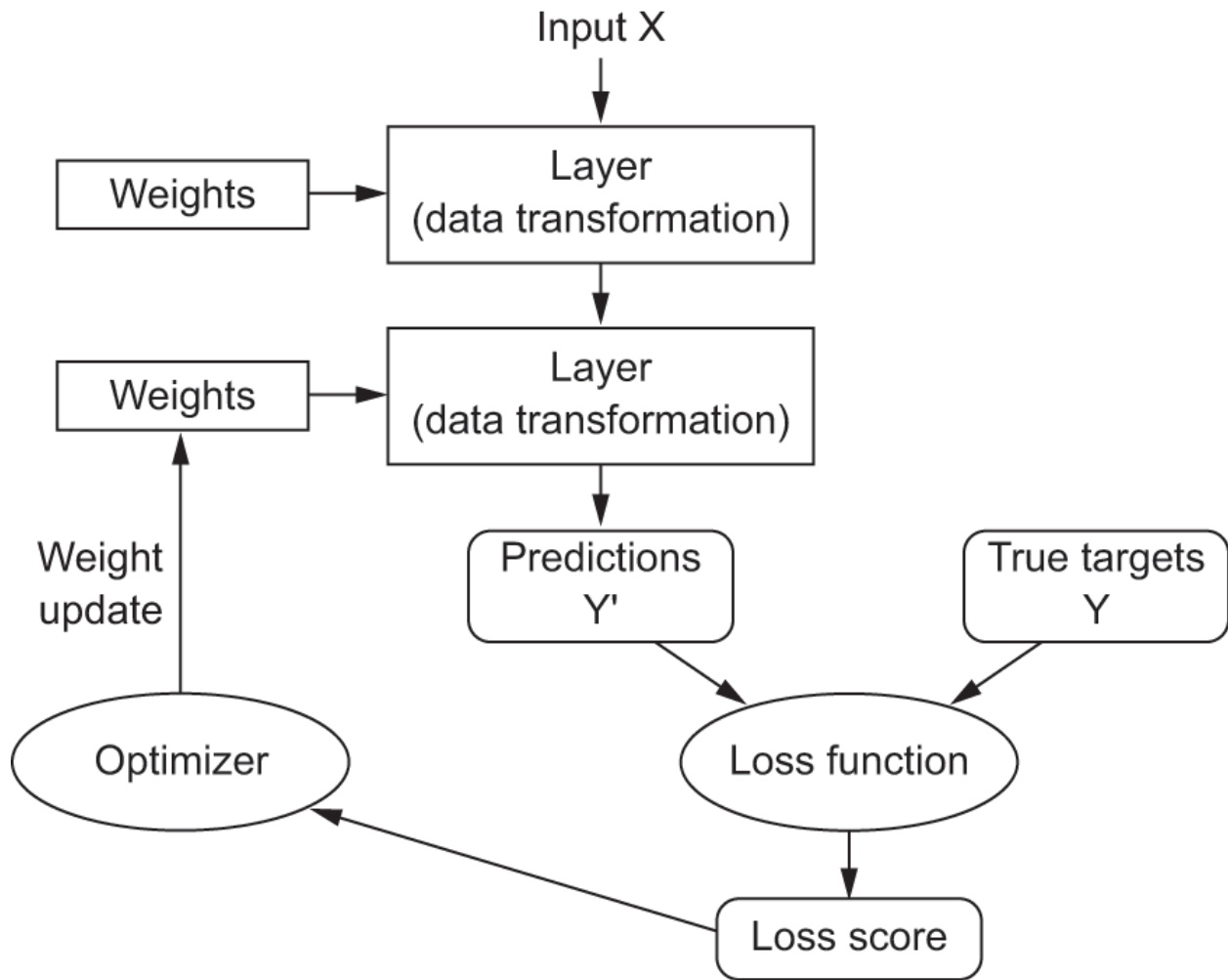
Εικόνα 22: Ένα TN παραμετροποιείται από τα βάρη του. [13]

Πώς όμως το TN καθορίζει ποιες τιμές θα αλλάξουν και πόσο; Κάθε φορά που το νευρωνικό δίκτυο κάνει μια πρόβλεψη, συγκρίνεται με τον πραγματικό στόχο. Ανάλογα με την απόσταση που απέχει από τον πραγματικό στόχο, υπολογίζεται μια τιμή. Αυτή η διαδικασία είναι δουλειά της συνάρτησης απώλειας (loss function) του δικτύου. [13]



Εικόνα 23: Η συνάρτηση απώλειας μετράει την ποιότητα της εξόδου του δικτύου [13]

Η τιμή που υπολογίζεται από την συνάρτηση απώλειας λέγεται Loss score. Αυτό το σκορ θα χρησιμοποιηθεί αργότερα ως σήμα ανάδρασης με σκοπό να προσαρμοστούν κατάλληλα η τιμές το βαρών, ώστε να μειωθεί. Αυτή η προσαρμογή είναι δουλειά του Βελτιστοποιητή (Optimizer) ο οποίος εφαρμόζει αυτό που ονομάζεται αλγόριθμος οπισθοδιάδοσης, ο οποίος είναι ο κεντρικός αλγόριθμος στην βαθιά μάθηση. [13]



Εικόνα 24: Το σκορ απώλειας χρησιμοποιείται σαν σήμα ανάδρασης για να προσαρμοστούν τα βάρη. [13]

Στην αρχή στα βάρη των δικτύων εκχωρούνται τυχαίες τιμές, κατά συνέπεια και το σκορ απώλειας είναι αρκετά υψηλό. Παρ' όλα αυτά, με κάθε παράδειγμα που το δίκτυο επεξεργάζεται, τα βάρη προσαρμόζονται καλύτερα και το σκορ απώλειας μειώνεται. Αυτός ο κύκλος εκπαίδευσης, ονομάζεται εποχή (epoch) και επαναλαμβάνεται αρκετές φορές μέχρι να αποδώσει τιμές στα βάρη που θα ελαχιστοποιήσουν την απώλεια. [13]

1.3.5 Τί έχει πετύχει η βαθιά μάθηση μέχρι σήμερα;

Παρόλο που η βαθιά μάθηση σαν υποπεδίο της μηχανικής μάθησης υπάρχει αρκετά χρόνια, ανέβηκε στο προσκήνιο στις αρχές του 2010. Μέσα σε λίγο χρόνο έχει πετύχει αξιοσημείωτα αποτελέσματα για προβλήματα σε διάφορους τομείς που περιλαμβάνουν δεξιότητες, απλές για τον άνθρωπο και μέχρι πρόσφατα δύσκολες για την μηχανή. [13]

Πιο συγκεκριμένα, η βαθιά μάθηση επέτρεψε τις ακόλουθες ανακαλύψεις, όλες σε ιστορικά δύσκολους τομείς για την μηχανική μάθηση: [13]

- Ταξινόμηση εικόνων σχεδόν σε ανθρώπινο επίπεδο.
- Μεταγραφή ομιλίας σχεδόν σε ανθρώπινο επίπεδο.
- Μεταγραφή χειρόγραφου σχεδόν σε ανθρώπινο επίπεδο.
- Βελτιωμένη αυτόματη μετάφραση.
- Βελτίωση στην μετατροπή κειμένου σε ομιλία.
- Ψηφιακοί βοηθοί όπως το Google Now και το Amazon Alexa.
- Αυτόνομη οδήγηση σχεδόν σε ανθρώπινο επίπεδο.
- Βελτιωμένη στόχευση διαφημίσεων.
- Βελτιωμένα αποτελέσματα αναζήτησης στον διαδίκτυο.
- Ικανότητα απάντησης σε ερωτήσεις φυσικής γλώσσας.
- Βελτίωση των AI χαρακτήρων – αντιπάλων στα βιντεοπαιχνίδια.

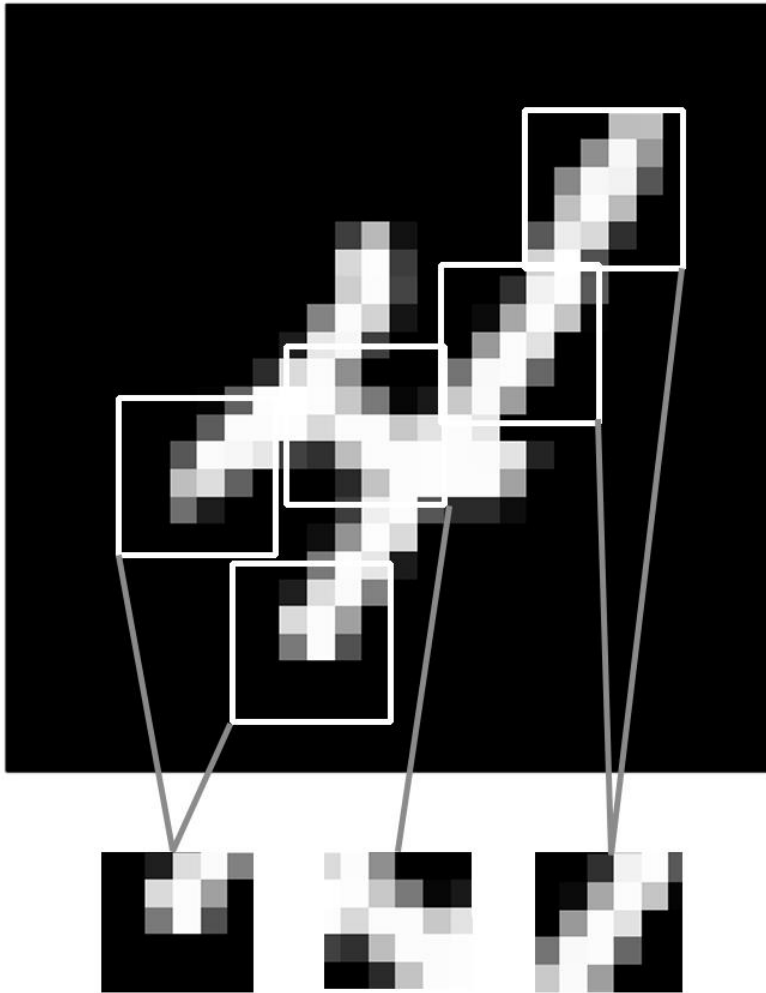
1.3.6 Βαθιά μάθηση για μηχανική όραση

Η μηχανική όραση είναι μία από τις πρώτες και μεγαλύτερες επιτυχίες της βαθιάς μάθησης. Χρήστες κινητών και σταθερών συσκευών έρχονται σε επαφή με βαθιά μοντέλα όρασης χωρία να το γνωρίζουν, μέσω φίλτρων βίντεο και φωτογραφιών σε λογισμικά κάμερας, στην αναζήτηση εικόνων μέσω Google, στο YouTube, σε εφαρμογές διαχείρισης εικόνων και πολλά άλλα. Επίσης τέτοια μοντέλα αποτελούν τον πυρήνα των ερευνών στην αυτόνομη οδήγηση, τη ρομποτική, την ιατρική, τα αυτόνομα συστήματα πληρωμών λιανικής, ακόμη και στην γεωργία.

Το πρόβλημα της μηχανική όραση λειτούργησε σαν παράγοντας για την αρχική άνοδο της βαθιάς μάθησης μεταξύ 2011 και 2015. Ένας τύπος βαθιών νευρωνικών δικτύων που ονομάζονται συνελκτικά νευρωνικά δίκτυα άρχισε να έχει εξαιρετικά αποτελέσματα σε διαγωνισμούς ταξινόμησης εικόνων εκείνη την εποχή, πρώτα με τον Ντάν Σιρεσάν (Dan Ciresan) να κερδίζει δύο διαγωνισμούς (τον διαγωνισμό αναγνώρισης κινεζικών χαρακτήρων ICDAR 2011 και τον διαγωνισμό αναγνώρισης σημάτων οδικής κυκλοφορίας IJCNN 2011), με αποκορύφωμα τον φθινόπωρο του 2012 η ομάδα του Χίλτον (Hilton) να κερδίζει την μεγάλη πρόκληση της οπτικής αναγνώρισης ImageNet. Από τότε εφαρμόστηκαν και σε άλλες εργασίες μηχανικής όρασης φέρνοντας πολλά υποσχόμενα αποτελέσματα. [13]

1.3.7 Η λειτουργία της συνέλιξης

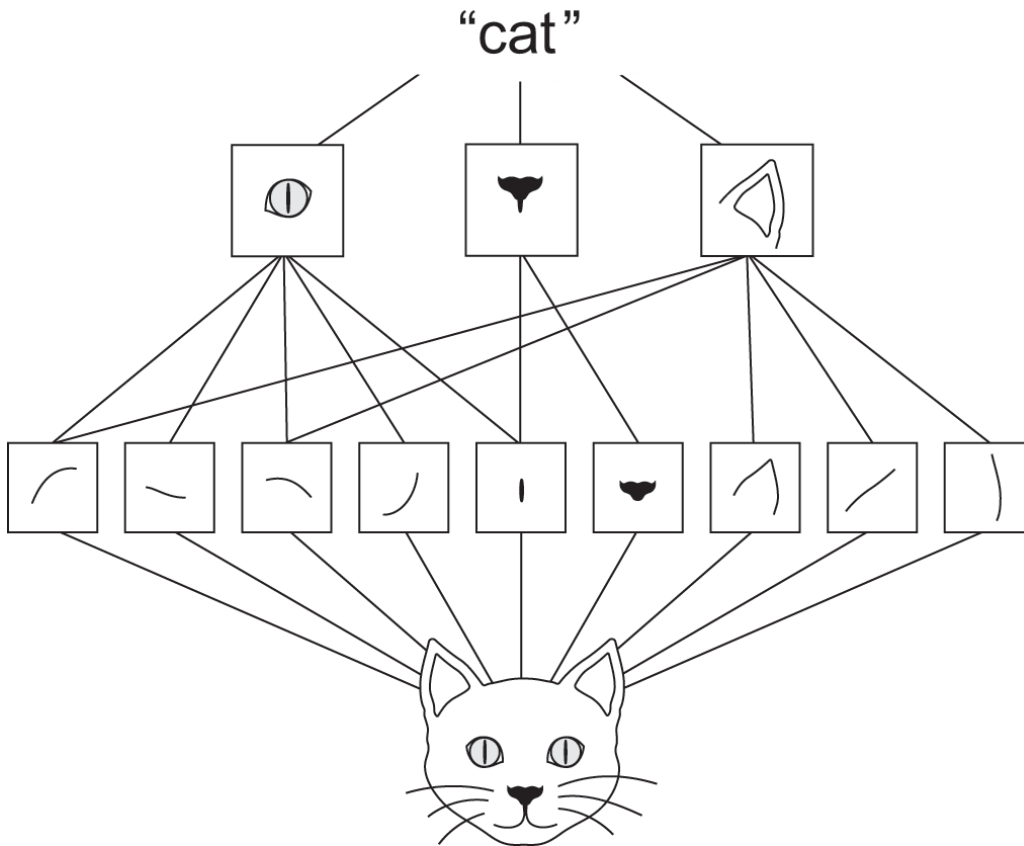
Η θεμελιώδης διαφορά μεταξύ ενός πυκνά συνδεδεμένου στρώματος (densely connected layer) και ενός στρώματος συνέλιξης (convolutional layer) είναι η εξής: Τα πυκνά στρώματα μαθαίνουν καθολικά μοτίβα στο χώρο των χαρακτηριστικών εισόδου τους (δηλαδή όλα τα εικονοστοιχεία), ενώ τα στρώματα συνέλιξης μαθαίνουν τοπικά μοτίβα, στην περίπτωση εικόνων, μοτίβα που βρίσκονται σε μικρά δυσδιάστατα παράθυρα των εισόδων.



Εικόνα 25: Οι εικόνες μπορούν να χωριστούν σε τοπικά μοτίβα, όπως ακμές, υφές και ούτω καθεξής. [13]

Αυτό το χαρακτηριστικό κλειδί δίνει τα συνελκτικά δίκτυα δύο ενδιαφέροντα χαρακτηριστικά:

- Τα πρότυπα που μαθαίνουν είναι αμετάβλητα στην μεταφορά (translation invariant). Αυτό σημαίνει ότι άπαξ και μάθει ένα μοτίβο σε οποιοδήποτε μέρος μιας εικόνας, θα μπορεί να το αναγνωρίσει οπουδήποτε. Αντίθετα ένα πυκνά συνδεδεμένο δίκτυο θα έπρεπε να το μάθει ξανά εάν εμφανιζόταν σε διαφορετική θέση. Κατά συνέπεια, τα συνελκτικά δίκτυα χρειάζονται λιγότερα δείγματα για να μάθουν γενικευμένες αναπαραστάσεις.
- Μπορούν να μάθουν χωρικές ιεραρχίες μοτίβων. Πιο αναλυτικά, τα πρώτα συνελκτικά στρώματα μαθαίνουν μικρά τοπικά μοτίβα όπως άκρα, το δεύτερα στρώματα θα μάθουν μεγαλύτερα μοτίβα φτιαγμένα από τα χαρακτηριστικά των πρώτων στρωμάτων και ούτω καθεξής. Αυτό επιτρέπει στα συνελκτικά δίκτυα να μαθαίνουν αποτελεσματικά όλο και πιο περίπλοκες και αφηρημένες οπτικές έννοιες. [13]



Εικόνα 26: Ο οπτικός κόσμος σχηματίζει μια χωρική ιεραρχία οπτικών εννοιών: στοιχειώδεις γραμμές ή υφές συνδυάζονται σε απλά αντικείμενα όπως μάτια ή αυτιά, τα οποία συνδυάζονται σε έννοιες υψηλού επιπέδου όπως η «γάτα». [13]

1.3.8 Τα τρία είδη εργασιών της Μηχανικής Όρασης

Υπάρχουν τρία διαφορετικά είδη εργασιών της μηχανικής όρασης που πρέπει να αναφερθούν. Αυτά τα είδη είναι η «κατηγοριοποίηση εικόνων», η «τμηματοποίηση εικόνων» και η «ανίχνευση αντικειμένων». Σ' αυτήν την εργασία χρησιμοποιήθηκαν τα πρώτα δύο. Πιο αναλυτικά:

- 1) Κατηγοριοποίηση Εικόνων (Image classification). Ο στόχος είναι να εκχωρηθούν μία ή περισσότερες ετικέτες σε μία εικόνα. Αυτό γιατί, υπάρχουν δύο είδη κατηγοριοποίησης εικόνων· η κατηγοριοποίηση μονής ετικέτας (μια εικόνα μπορεί να ανήκει μόνο σε μία κατηγορία) και πολλαπλών ετικετών (όπου μια εικόνα μπορεί να ανήκει σε παραπάνω από μια κατηγορία). Για παράδειγμα, όταν αναζητάτε μια λέξη-κλειδί στην εφαρμογή Google Photos, στα παρασκήνια ρωτάτε ένα πολύ μεγάλο μοντέλο ταξινόμησης πολλών ετικετών με περισσότερες από 20.000 διαφορετικές κατηγορίες, εκπαιδευμένο σε εκατομμύρια εικόνες.
- 2) Τμηματοποίηση Εικόνων (Image segmentation). Όπου ο στόχος είναι "η τμηματοποίηση" ή "ο διαμερισμός" μιας εικόνας σε διαφορετικές περιοχές, με κάθε περιοχή να αντιπροσωπεύει συνήθως μια κατηγορία. Για παράδειγμα τα προσαρμοσμένα φόντο που έχουν για βιντεοκλήσεις λογισμικά σαν το Microsoft Teams, χρησιμοποιούν ένα μοντέλο τμηματοποίησης εικόνας για να ξεχωρίζουν το πρόσωπό σας και το τι υπάρχει πίσω του, με ακρίβεια pixel.
- 3) Ανίχνευση Αντικειμένων (Object detection). Όπου ο στόχος είναι να σχεδιαστούν ορθογώνια (δηλαδή πλαίσια οριοθέτησης) γύρω από αντικείμενα ενδιαφέροντος σε μια εικόνα και να συσχετιστεί κάθε ορθογώνιο με μια κατηγορία. Για παράδειγμα, ένα αυτόνομο αυτοκίνητο θα μπορούσε να χρησιμοποιήσει ένα μοντέλο ανίχνευσης αντικειμένων για την παρακολούθηση αυτοκινήτων, πεζών και πινακίδων ενόψει των καμερών του.

Στην πραγματικότητα, η βαθιά μάθηση για την μηχανική όραση περιλαμβάνει και άλλες πιο εξειδικευμένες εργασίες εκτός από αυτές τις τρεις. Μερικές από αυτές είναι η βαθμολόγηση της ομοιότητα εικόνων, ανίχνευση σημείων-κλειδιών (εντοπισμός χαρακτηριστικών ενδιαφέροντος σε μια εικόνα, όπως χαρακτηριστικά προσώπου), εκτίμηση στάσης σώματος, εκτίμηση πλέγματος τριών διαστάσεων (3D) και ούτω καθεξής. Ωστόσο οι τρεις βασικές που αναφέρθηκαν παραπάνω αποτελούν τα θεμέλια της μηχανικής όρασης.

Κεφάλαιο 2: Μέθοδος Επίλυσης

Σε αυτό το κεφάλαιο θα παρουσιαστεί βήμα-βήμα η διαδικασία που ακολουθήθηκε ώστε να λυθεί το πρόβλημα.

2.1 Tensorflow & Keras

Το TensorFlow είναι μια ολοκληρωμένη πλατφόρμα ανοιχτού κώδικα για μηχανική μάθηση την οποία ανέπτυξε η Google Brain, ομάδα τεχνητής νοημοσύνης της Google, αρχικά για εσωτερική χρήση. Διαθέτει ένα ολοκληρωμένο, ευέλικτο οικοσύστημα εργαλείων, βιβλιοθηκών και πόρων της κοινότητάς του, που επιτρέπει στους ερευνητές να εξελίσσουν την τελευταία λέξη της τεχνολογίας στη Μηχανική Μάθηση και οι προγραμματιστές να δημιουργούν και να αναπτύσσουν εύκολα εφαρμογές που υποστηρίζονται από τη Μηχανική Μάθηση. Μπορεί να χρησιμοποιηθεί σε ένα ευρύ φάσμα εργασιών, αλλά έχει ιδιαίτερη έμφαση στην εκπαίδευση και την εξαγωγή συμπερασμάτων χρησιμοποιώντας βαθιά νευρωνικά δίκτυα. [14] [15]

Το Keras είναι μια βιβλιοθήκη λογισμικού ανοιχτού κώδικα που παρέχει μια διεπαφή Python για τεχνητά νευρωνικά δίκτυα. Το Keras λειτουργεί ως διεπαφή για τη πλατφόρμα TensorFlow. Το Keras ακολουθεί τις βέλτιστες πρακτικές για τη μείωση του γνωστικού φορτίου: προσφέρει συνεπή και απλά API, ελαχιστοποιεί τον αριθμό των ενεργειών του χρήστη που απαιτούνται για συνήθεις περιπτώσεις και παρέχει σαφή και διαδραστικά μηνύματα σφαλμάτων. Διαθέτει επίσης εκτενή τεκμηρίωση και οδηγούς προγραμματιστών. [16] [17]

2.2 Γενίκευση και Βελτιστοποίηση

Το βασικό ζήτημα στη μηχανική μάθηση είναι η ένταση μεταξύ βελτιστοποίησης και γενίκευσης. Η βελτιστοποίηση αναφέρεται στη διαδικασία προσαρμογής ενός μοντέλου για την καλύτερη δυνατή απόδοση στα δεδομένα εκπαίδευσης, ενώ η γενίκευση αναφέρεται στο πόσο καλά αποδίδει το εκπαιδευμένο μοντέλο σε δεδομένα που δεν έχει ξαναδεί. Οπότε όπως είναι φυσικό ο σκοπός είναι να πετύχουμε καλή γενίκευση, κάτι όμως που δεν περνάει εξ' ολοκλήρου από το χέρι του μηχανικού. Το μόνο που μπορεί να κάνει ένας μηχανικός είναι να εκπαιδεύσει το μοντέλο του πάνω στα δεδομένα που έχει, αν το κάνει πολύ καλά (σε ακραίο βαθμό) η γενίκευση θα υποφέρει. [13]

2.3 Υπερπροσαρμογής (Overfitting) & Υποπροσαρμογής (Underfitting)

Προτού παρουσιασθεί η διαδικασία που ακολουθήθηκε, κρίνεται αναγκαίο να εξηγηθούν οι όροι «overfitting» και «underfitting», μιας και ακούγονται συχνά στον κόσμο της Μηχανικής Μάθησης. Το «underfitting» είναι ένα σενάριο στο οποίο το μοντέλο είναι ανίκανο να μάθει τα σχετικά μοτίβα στα δεδομένα εκπαίδευσης. Το «overfitting» είναι το ακριβώς αντίθετο, δηλαδή

μαθαίνει μοτίβα ειδικά για τα δεδομένων εκπαίδευσης τα οποία όμως είναι άσχετα όταν πρόκειται για νέα δεδομένα.

2.3.1 Τρόποι καταπολέμησης της Υπερπροσαρμογής (Overfitting) & Υποπροσαρμογής (Underfitting)

Μερικοί από τους τρόπους μείωσης του «overfitting» είναι οι εξής:

- Μείωση του μεγέθους του δικτύου: Ο πιο απλός τρόπος για να μειωθεί η υπερπροσαρμογή είναι η μείωση του μεγέθους του δικτύου, αυτό γιατί ο αριθμός των παραμέτρων που μπορεί να μάθει, καθορίζεται από τον αριθμό των στρωμάτων και των μονάδων ανά επίπεδο. Εάν το μοντέλο έχει περιορισμένους πόρους απομνημόνευσης, δεν θα είναι σε θέση να απομνημονεύσει τα δεδομένα εκπαίδευσης. Έτσι, για να ελαχιστοποιήσει την απώλειά του, θα πρέπει να καταφύγει στη μάθηση συμπίεσμένων παραστάσεων που έχουν προγνωστική ισχύ ως προς τους στόχους. Ωστόσο αν ο αριθμός των παραμέτρων μειωθεί σε υπερβολικό βαθμό, μπορεί να εμφανιστεί το πρόβλημα της υποπροσαρμογής όπου το μοντέλο αδυνατεί να μάθει το στιδήποτε.
- Εφαρμογή κανονικοποίησης βάρους (weight regularization): Ένας κοινός τρόπος για να μετριαστεί η υπερπροσαρμογή (overfitting) είναι να τεθούν περιορισμοί στην πολυπλοκότητα ενός μοντέλου αναγκάζοντας τα βάρη του να λάβουν μόνο μικρές τιμές, γεγονός που καθιστά την κατανομή των τιμών των βαρών πιο ομαλή. Αυτό επιτυγχάνεται προσθέτοντας στη συνάρτηση απώλειας του μοντέλου ένα κόστος που σχετίζεται με την ύπαρξη μεγάλων βαρών. Υπάρχουν δύο είδη κανονικοποίησης βάρους.
 - Κανονικοποίηση L1 (L1 regularization): Το προστιθέμενο κόστος είναι ανάλογο με την απόλυτη τιμή των συντελεστών βάρους.
 - Κανονικοποίηση L2 (L2 regularization): Το προστιθέμενο κόστος είναι ανάλογο με το τετράγωνο της αξίας των συντελεστών βάρους. Η κανονικοποίηση L2 είναι γνωστή και με το όνομα φθίνων βάρους (weight decay).
- Χρήση του Dropout layer: Το Dropout είναι μια από τις πιο αποτελεσματικές και συχνότερα χρησιμοποιούμενες τεχνικές κανονικοποίησης για νευρωνικά δίκτυα, που αναπτύχθηκε από τον Geoff Hinton και τους μαθητές του στο Πανεπιστήμιο του Τορόντο. Το Dropout, που εφαρμόζεται σε ένα επίπεδο, αποτελείται από την τυχαία εγκατάλειψη (ρύθμιση στο μηδέν) ορισμένων χαρακτηριστικών εξόδου των επιπέδων κατά τη διάρκεια της εκπαίδευσης. Η βασική ιδέα είναι ότι η εισαγωγή θορύβου στις τιμές εξόδου ενός στρώματος μπορεί να διασπάσει τυχαία μοτίβα που δεν είναι σημαντικά, τα οποία το μοντέλο θα απομνημόνευε εάν δεν υπήρχε θόρυβος.
- Επιμέλεια συνόλου δεδομένων: Υπάρχει ένας γενικός κανόνας στη μηχανική μάθηση που περιγράφεται από τη φράση «ό,τι δίνεις παίρνεις» ή όπως λένε οι Άγγλοι “garbage in garbage out”. Όσο καλό και καλά ρυθμισμένο να είναι ένα μοντέλο αν τα σύνολο δεδομένων δεν είναι καλό τότε και τα αποτελέσματα δεν θα είναι εξίσου καλά. Στην

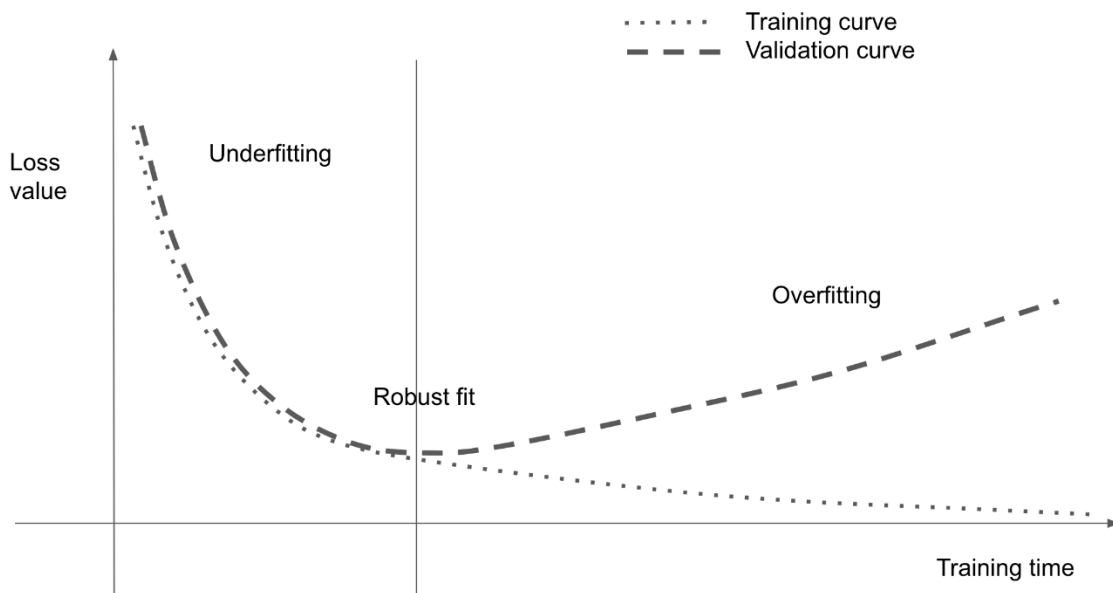
τελική η βαθιά μάθηση προσαρμόζει καμπύλες, δεν κάνει μαγικά! Τα βήματα που μπορεί να ακολουθήσει κάποιος είναι τα εξής:

- Τα δεδομένα πρέπει να είναι αρκετά ώστε να εκπαιδευτεί καλά το μοντέλο. Σε περίπτωση που δεν υπάρχει δυνατότητα εύρεσης νέων, μπορούν να δημιουργηθούν νέα, μέσω μιας διαδικασίας που λέγεται αύξηση δεδομένων (augmentation).
- Επίσης καλό είναι να περιοριστούν τα λάθη επισήμανσης (labeling errors). Για παράδειγμα ένα λάθος επισήμανσης είναι μια εικόνα γάτας με την ετικέτα «σκύλος».
- Μια καλή τεχνική είναι και ο καθαρισμός των δεδομένων από ελλείπουσες τιμές και τιμές που δεν σχετίζονται με τον στόχο. Ουκ ολίγες φορές ένα σύνολο δεδομένων μπορεί να έχει μη συμπληρωμένες τιμές καθώς και άχρηστα χαρακτηριστικά που δεν παρέχουν πληροφορίες για τον στόχο. Ο μηχανικός οφείλει να εξερευνήσει το σύνολο των δεδομένων και να αντιμετωπίσει τέτοια προβλήματα.
- Μηχανική Χαρακτηριστικών (features engineering): Το “feature engineering” είναι η διαδικασία χρήσης των γνώσεων του μηχανικού σχετικά με τα δεδομένα και με τον αλγόριθμο μηχανικής εκμάθησης που θα εφαρμοστεί, για να τον κάνει να λειτουργεί καλύτερα εφαρμόζοντας μετασχηματισμένους στα δεδομένα πριν προχωρήσουν στο μοντέλο. Με λίγα λόγια η ουσία της μηχανικής χαρακτηριστικών είναι, η μετατροπή ενός προβλήματος σε πιο απλό μέσω της απλούστερης έκφρασής του. Αυτό συνήθως απαιτεί κατανόηση του προβλήματος σε βάθος. Πλέον, η σύγχρονη βαθιά μάθηση δεν έχει τόσο ανάγκη την μηχανική χαρακτηριστικών, επειδή τα νευρωνικά δίκτυα είναι σε θέση να εξάγουν αυτόματα χρήσιμες λειτουργίες από ακατέργαστα δεδομένα. Αυτό όμως δεν σημαίνει ότι δεν χρειάζεται να χρησιμοποιείται η μηχανική χαρακτηριστικών, για δύο λόγους:
 - Τα καλά χαρακτηριστικά εξακολουθούν να επιτρέπουν την επίλυση προβλημάτων με κομψότερο τρόπο ενώ χρησιμοποιούνται λιγότεροι πόροι.
 - Τα καλά χαρακτηριστικά επιτρέπουν την λύση προβλημάτων με τη βοήθεια πολύ λιγότερων δεδομένων. Η ικανότητα των μοντέλων βαθιάς εκμάθησης να μαθαίνουν μόνοι τους χαρακτηριστικά βασίζεται στο να υπάρχουν πολλά διαθέσιμα δεδομένα. Άρα αν ο αριθμός των δειγμάτων είναι περιορισμένος τότε η αξία των πληροφοριών στα χαρακτηριστικά τους καθίσταται κρίσιμη.
- Χρήση της πρόωρης διακοπής (Early Stopping): Ένας άλλος τρόπος καταπολέμησης της υπερπροσαρμογής είναι μέσω της πρόληψης, δηλαδή να σταματήσει η εκπαίδευση πριν το μοντέλο φτάσει σε αυτό το σημείο. Αυτό επίσης δίνει και την ελευθερία στον μηχανικό να φτιάξει ένα υπερπαραμετροποιημένο μοντέλο χωρίς να φοβάται για την υπερπροσαρμογή. Το callback “Early Stopping” της βιβλιοθήκης Keras θα διακόψει την διαδικασία της εκμάθησης μόλις σταματήσει να βελτιώνεται η παρακολουθούμενη μετρούμενη μονάδα.

Μερικοί από τους τρόπους μείωσης του «underfitting» είναι οι εξής:

- Λήψη ή δημιουργία περισσότερων δεδομένων και μηχανική χαρακτηριστικών. Όπως αναφέρθηκε και παραπάνω η ύπαρξη ενός συνόλου δεδομένων με αρκετά δεδομένα είναι ουσιώδης. Αν δεν υπάρχει η δυνατότητα εύρεσης νέων μπορούν να δημιουργηθούν νέα εφαρμόζοντας μετασχηματισμούς τα υπάρχοντα μέσω τεχνικών αύξησης δεδομένων. Επίσης όπως αναφέρθηκε και παραπάνω η μηχανική χαρακτηριστικών επιτρέπει τη λύση προβλημάτων με τη βοήθεια πολύ λιγότερων δεδομένων.
- Αύξηση του μεγέθους ή του αριθμού των παραμέτρων του μοντέλου. Επειδή ο αριθμός των παραμέτρων που μπορεί να μάθει ένα δίκτυο, καθορίζεται από τον αριθμό των στρωμάτων και των μονάδων ανά επίπεδο, αν αυξηθούν ένα από τα δύο ή και τα δύο, το δίκτυο θα είναι εξοπλισμένο με περισσότερους πόρους απομνημόνευσης.
- Αύξηση της πολυπλοκότητας του μοντέλου. Η «πολυπλοκότητα του μοντέλου» αναφέρεται συνήθως στον αριθμό των βαθμών ελευθερίας ενός μοντέλου, που συχνά μετράται ως ο αριθμός των ρυθμιζόμενων βαρών ή παραμέτρων στην αρχιτεκτονική του. Αυξάνοντας όμως την πολυπλοκότητα του μοντέλου έχοντας μικρό αριθμό δεδομένων μπορεί να προκαλέσει υπερπροσαρμογή στα δεδομένα εκπαίδευσης, οπότε θέλει προσοχή.
- Αύξηση του χρόνου εκπαίδευσης, έως ότου ελαχιστοποιηθεί η απώλεια (log loss). Αρκετές φορές τα δίκτυα χρειάζονται αρκετές εποχές για να αρχίσουν να μαθαίνουν. Με τη χρήση του callback πρόωρης διακοπής, ο αριθμός των εποχών μπορεί να πάρει μεγάλη τιμή χωρίς να υπάρχει ο φόβος της υπερπροσαρμογής.

Ακολουθώντας τις παραπάνω τεχνικές, θα μπορούσε κανείς να βελτιώσει την απόδοση των μοντέλων του και να αντιμετωπίσει τα προβλήματα του overfitting και underfitting. [18] [13]

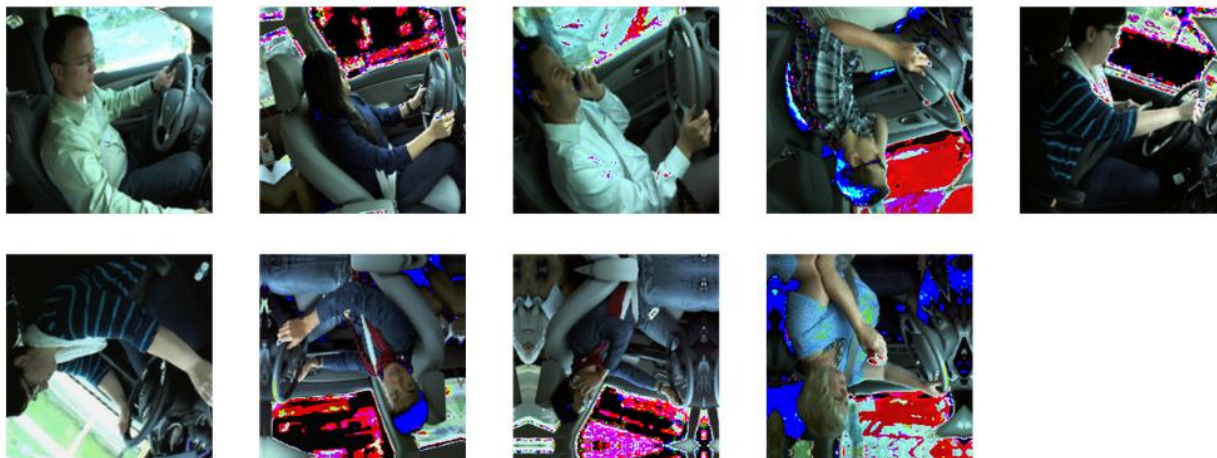


Εικόνα 27: Κανονική Overfitting συμπεριφορά. [13]

2.3.2 Αύξηση δεδομένων (Data Augmentation)

Όπως έχει προαναφερθεί ο αριθμός των δεδομένων είναι μόνο 22.423 φωτογραφίες. Μπορεί αυτός ο αριθμός να φαίνεται μεγάλος, άλλα για την εκπαίδευση νευρωνικού δικτύου είναι μικρός. Χρησιμοποιώντας την βιβλιοθήκη Keras - preprocessing του Tensorflow, θα δημιουργηθούν νέα δεδομένα τροποποιώντας τα υπάρχοντα. Πιο συγκεκριμένα, αυτές οι τροποποιήσεις είναι:

- Τυχαία αναστροφή, οριζόντια και κάθετα (Random Flip). Τυχαία αναστροφή κάθε εικόνας οριζόντια και κάθετα.
- Τυχαία περιστροφή (Random Rotation). Τυχαία περιστροφή κάθε εικόνας κατά την εκπαίδευση.
- Τυχαία μεγέθυνση (Random Zoom). Τυχαία μεγέθυνση κάθε εικόνας κατά τη διάρκεια της εκπαίδευσης.
- Τυχαία αντίθεση (Random Contrast). Προσαρμογή της αντίθεσης μιας εικόνας με έναν τυχαίο παράγοντα.



Εικόνα 28: Παραδείγματα τροποποιημένων εικόνων

Υπάρχουν όμως κι άλλες τροποποιήσεις που δεν επιλέχθηκαν, όπως η:

- Τυχαία περικοπή (Random Crop). Τυχαία περικοπή των εικόνων στο δοσμένο ύψος και πλάτος
- Τυχαία μεταφορά (Random Translation). Εφαρμόζει τυχαίες μεταφορές σε κάθε εικόνα κατά την εκπαίδευση.
- Τυχαίο ύψος (Random Height). Τυχαία μεταβολή του ύψους μιας παρτίδας εικόνων κατά τη διάρκεια της εκπαίδευσης.
- Τυχαίο πλάτος (Random Width). Τυχαία μεταβολή του πλάτους μιας παρτίδας εικόνων κατά τη διάρκεια της εκπαίδευσης.

Εκτός από το keras-preprocessing υπάρχει και άλλη μέθοδος για αύξηση των δεδομένων και λέγεται ImageDataGenerator. Η εν λόγω μέθοδος πέρα από τις τεχνικές αύξησης δεδομένων που θα αναφερθούν παρακάτω, έχει τεχνικές κανονικοποίησης (όπως η ανακλιμάκωση που επαναπροσδιορίζει και αντισταθμίζει τις τιμές μιας εικόνας), καθώς και άλλες ρυθμίσεις που δεν θα αναφερθούν γιατί δεν είναι της παρούσης.

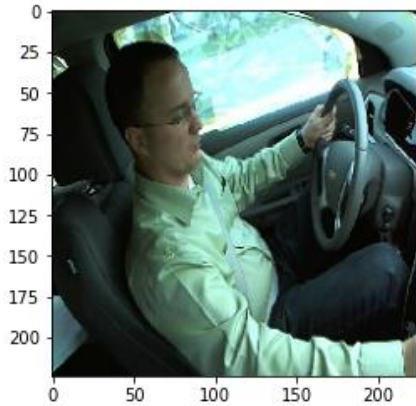
- Εύρος περιστροφής (rotation range). Δέχεται ακέραιο αριθμό που ρυθμίζει το εύρος βαθμού για τυχαίες περιστροφές.
- Εύρος φωτεινότητας (brightness range). Εύρος για επιλογή τιμής μετατόπισης φωτεινότητας.
- Εύρος διάτμησης (shear range). Ένταση διάτμησης (γωνία διάτμησης αριστερόστροφα σε μοίρες).
- Εύρος μεγέθυνσης (zoom range). Εύρος για τυχαία μεγέθυνση.
- Οριζόντια αναστροφή (horizontal flip). Τυχαία περιστροφή των εικόνων οριζόντια.
- Κάθετη αναστροφή (vertical flip). Τυχαία περιστροφή των εικόνων κάθετα.
- Εύρος μετατόπισης πλάτους (width shift range). Μετατόπιση των εικόνων προς τα αριστερά ή τα δεξιά
- Εύρος μετατόπισης ύψους (height shift range). Μετατόπιση των εικόνων προς τα πάνω ή κάτω.

2.3.3 Θόλωση προσώπων

Κατά την εξερεύνηση των δεδομένων του State Farm Dataset στο Κεφάλαιο 1, ένα από τα προβλήματα που επισημάνθηκαν ήταν η εμφάνιση κάποιων οδηγών περισσότερο από κάποιους άλλους. Εκτός από τις δυσκολίες που μπορεί προκαλέσει αυτό, τίθεται επίσης και το θέμα της προστασίας των προσωπικών δεδομένων των οδηγών. Προφανώς οι οδηγοί που εμφανίζονται στις φωτογραφίες του dataset έχουν συμφωνήσει για κάτι τέτοιο, ωστόσο στην πραγματικότητα όταν το σύστημα του κάθε αυτοκινήτου θα συλλέγει δεδομένα για την βελτίωση της απόδοσης του μοντέλου, δεν θα υπάρχει τέτοια δυνατότητα.

Για την επίλυση αυτού του προβλήματος, δύο εκδοχές εξερευνήθηκαν. Η πρώτη ήταν μια συνάρτηση (function) προεπεξεργασίας, η οποία χρησιμοποιούσε έναν ταξινομητή Haar-Cascade της βιβλιοθήκης OpenCV για τον εντοπισμό και την θόλωση των προσώπων των οδηγών. Η δεύτερη ήταν μια function προεπεξεργασίας, η οποία εφάρμοζε θόλωση σε μια περιοχή ενδιαφέροντος (ROI - region of interest) η οποία είχε οριστεί μέσω κώδικα.

Τελικά από τις δύο επικράτησε η δεύτερη γιατί η πρώτη απαιτούσε παραπάνω επεξεργαστική ισχύ και πόρους που υπήρχαν προ διάθεση. Επίσης καθυστερούσε περισσότερο την διαδικασία της εκπαίδευσης. Παρ' όλα αυτά, αν υπήρχε καλύτερος εξοπλισμός, ο πρώτος τρόπος θα είχε προτιμηθεί.



Εικόνα 29: Εικόνα οδηγού χωρίς θόλωση προσώπου



Εικόνα 30: Εικόνα οδηγού με θόλωση προσώπου

2.3.4 Βάρη Κατηγοριών (Class Weights)

Το δεύτερο πρόβλημα που επισημάνθηκε στο Κεφάλαιο 1 ήταν η ταξική ανισορροπία, δηλαδή κάποιες κατηγορίες περιείχαν περισσότερα δεδομένα από κάποιες άλλες. Ο τρόπος που λύθηκε το πρόβλημα είναι μέσω της μεθόδου «κατηγορία βαρών» (class weights), όπου γίνεται επιλογή της κατηγορίας με τα περισσότερα δεδομένα ως αναφορά και την εκχώρηση των βαρών για τις άλλες κατηγορίες με βάση την κατηγορία αναφοράς. Για παράδειγμα αν έχουμε 3 κατηγορίες A: 10%, B: 50% και C: 40%, τότε τα βάρη γίνονται: {0:5, 1:1, 2:1.25}

Αυτό σημαίνει ότι αν το μοντέλο κατηγοριοποιήσει λανθασμένα την κατηγορία A, τότε η απώλεια (loss) θα είναι 5 φορές μεγαλύτερη από το να κατηγοριοποιήσει λανθασμένα την κατηγορία B.

Υπάρχει και άλλη μια μέθοδος, που δεν προτιμήθηκε, αυτή της «υπερβολικής δειγματοληψία» (Oversampling). Μια σχετική προσέγγιση θα ήταν η επαναδειγματοληψία του συνόλου δεδομένων αυξάνοντας τα δείγματα των κατηγοριών με τα λιγότερα δείγματα.

2.3.5 Τμηματοποίηση Εικόνας (Image Segmentation)

Ένας άλλος τρόπος για να βελτιώσουμε την ποιότητα των δεδομένων είναι μέσω μιας διαδικασίας που λέγεται Τμηματοποίηση Εικόνας (Image Segmentation). Έστω ότι χρειάζεται να βρεθεί η τοποθεσία ενός ή περισσότερων αντικειμένων σε μια εικόνα, δηλαδή ποια pixel ανήκουν στο εκάστοτε αντικείμενο, σε αυτήν την περίπτωση πρέπει να τμηματοποιηθεί (segment) η εικόνα, ήτοι σε κάθε pixel να δοθεί μια ετικέτα. Εξού και το καθήκον της τμηματοποίησης εικόνας (image segmentation) να εκπαιδεύσει ένα νευρωνικό δίκτυο ώστε να παράγει μια μάσκα εικονοστοιχείων (pixel-wise mask) για κάθε φωτογραφία. Αυτό βοηθά στην βαθύτερη κατανόηση της εικόνας στο επίπεδο των pixel. Αυτή η μέθοδος έχει πολλές εφαρμογές, όπως στην ιατρική, δορυφορική απεικόνιση και στα αυτοκινούμενα οχήματα. [19]

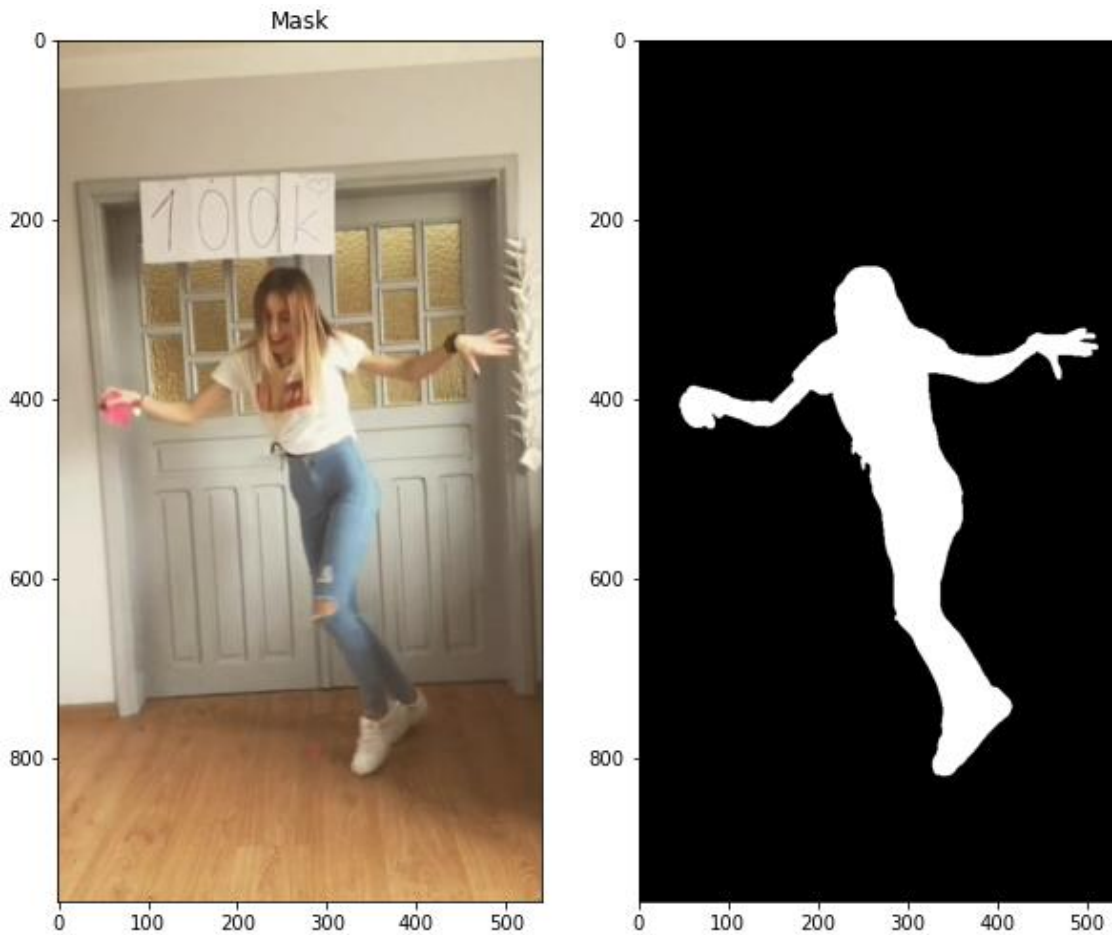
Υπάρχουν δύο διαφορετικές είδη τμηματοποίησης εικόνας:

- Σημασιολογική τμηματοποίηση (Semantic segmentation), όπου κάθε εικονοστοιχείο ταξινομείται ανεξάρτητα σε μια σημασιολογική κατηγορία, όπως "γάτα". Εάν υπάρχουν δύο γάτες στην εικόνα, τα αντίστοιχα εικονοστοιχεία αντιστοιχίζονται όλα στην ίδια γενική κατηγορία της "γάτας".
- Τμηματοποίηση παρουσίας (Instance segmentation), η οποία επιδιώκει όχι μόνο την ταξινόμηση των εικονοστοιχείων εικόνας ανά κατηγορία, αλλά και τον διαχωρισμό μεμονωμένων εμφανίσεων αντικειμένων. Σε μια εικόνα με δύο γάτες, η τμηματοποίηση παρουσίας θα αντιμετωπίζει το "γάτα 1" και το "γάτα 2" ως δύο ξεχωριστές κατηγορίες εικονοστοιχείων. [13]



Εικόνα 31: Σημασιολογική τμηματοποίηση και Τμηματοποίηση παρουσίας. [13]

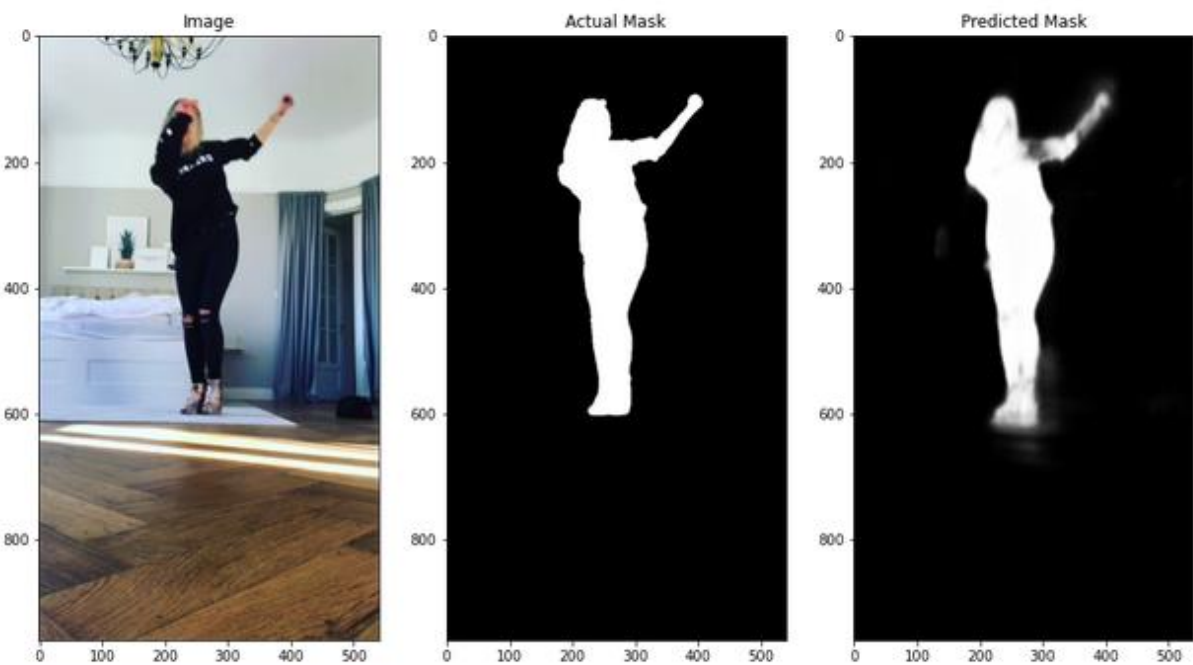
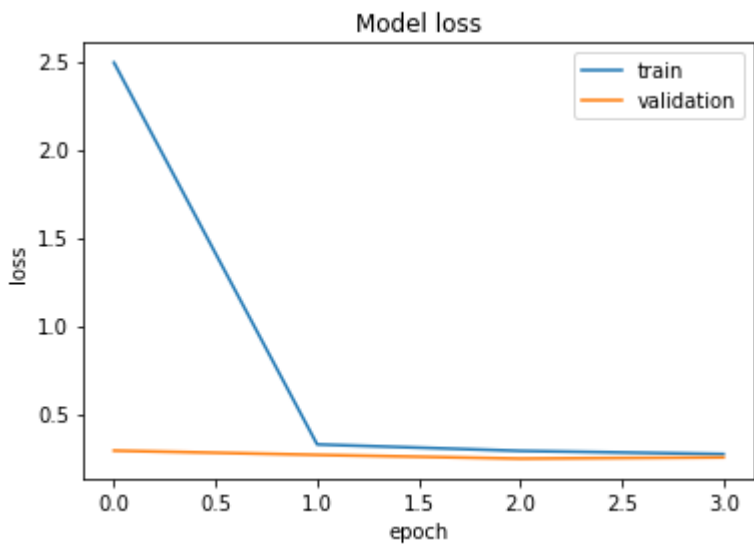
Το είδος τμηματοποίησης που επιλέχθηκε είναι η σημασιολογική τμηματοποίηση και πιο συγκεκριμένα αυτή του σώματος (Body Segmentation). Το πρώτο βήμα είναι να βρεθεί ένα σύνολο δεδομένων με εικόνες και τις μάσκες τους. Δυστυχώς δεν βρέθηκε κάποιο σχετικό σύνολο δεδομένων, δηλαδή με εικόνες οδηγών, οπότε χρησιμοποιήθηκε ένα με εικόνες χρηστών της πλατφόρμας του TikTok που χόρευαν μπροστά στην κάμερα. Το όνομα του είναι «Segmentation Full Body TikTok Dancing Dataset» και βρέθηκε στο site Kaggle.



Εικόνα 32: Μια από τις εικόνες μαζί με την αντίστοιχη μάσκα της

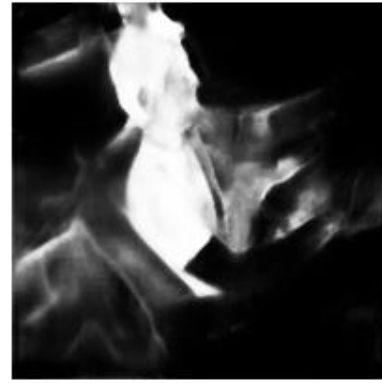
Η αρχιτεκτονική το μοντέλου που χρησιμοποιήθηκε είναι τύπου U-Net. Το U-Net είναι ένα convolutional νευρωνικό δίκτυο που αναπτύχθηκε για τμηματοποίηση εικόνων στον τομέα της βιοϊατρικής στο Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου του Φράμπουργκ (Freiburg). Το δίκτυο βασίζεται σε ένα πλήρως convolutional δίκτυο και η αρχιτεκτονική του τροποποιήθηκε και επεκτάθηκε ώστε να λειτουργεί με λιγότερες εκπαιδευτικές εικόνες και να δίνει ακριβέστερες τμηματοποιήσεις. [20]

Για την εκπαίδευση του χρησιμοποιήθηκε ο βελτιστοποιητής (Optimizer) Adam, για learning rate ορίστηκε η τιμή 10^{-3} και στο batch size η τιμή 16. Το μοντέλο εκπαιδεύτηκε για 28 εποχές και έφερε τα εξής αποτελέσματα: loss: 0.0518 - accuracy: 0.9711 - val_loss: 0.0684 - val_accuracy: 0.9667.



Εικόνα 34: Μια πρόβλεψη τυχαία επιλεγμένης εικόνας

Τέλος χρησιμοποιήθηκε το παραπάνω μοντέλο για να τμηματοποιήσει ολόκληρο το σύνολο δεδομένων του State Farm. Δυστυχώς όμως το αποτέλεσμα δεν ήταν το αναμενόμενο. Κάποιες εικόνες τμηματοποιήθηκαν ικανοποιητικά ενώ κάποιες άλλες όχι.



Εικόνα 35: Παράδειγμα καλής τμηματοποίησης.



Εικόνα 36: Παράδειγμα κακής τμηματοποίησης.

2.4 Σχετικές Εργασίες

Στην αρχή της ενασχόλησης με την εν λόγω εργασία, για την απόκτηση μιας πιο εμπειρισταωμένης γνώσης του αντικειμένου, μελέτησα τα παρακάτω επιστημονικά άρθρα.

- Distracted Driver Detection and Classification (Prof. Pramila M. Chawan, Shreyas Satardekar, Dharmin Shah, Rohit Badugu, Abhishek Pawar). Σ' αυτό το άρθρο δοκίμασαν διάφορα συνελκτικά μοντέλα, το καλύτερό τους όμως ήταν ένα ensemble που δημιουργήθηκε παίρνοντας τον μέσο όρο των πιθανοτήτων που δημιουργήθηκαν από τα μοντέλα VGG-16, VGG-19 και InceptionV3. Η τελική απώλεια ήταν 0.795.
- Distracted Driver Detection: Deep Learning vs Handcrafted Features (Murtadha D Hssayeni, Sagar Saxena, Raymond Ptucha, Andreas Savakis). Σε αυτήν την περίπτωση σύγκριναν επεξεργασμένα χαρακτηριστικά σε συνδυασμό με έναν ταξινομητή Support Machine Vector, με 3 βαθιά συνελκτικά νευρωνικά δίκτυα. Τα επεξεργασμένα χαρακτηριστικά περιλάμβαναν ένα μείγμα ιστογράμματος προσανατολισμένων διαβαθμίσεων (histogram of oriented gradients) και έναν προσδιοριστή κλιμακωτά

αμετάβλητου μετασχηματισμού χαρακτηριστικών (Scale-Invariant Feature Transform) που χρησιμοποιούνται για τη δημιουργία ενός Bag of Words μοντέλου. Τα τρία συνελκτικά δίκτυα που χρησιμοποιήθηκαν με την μέθοδο του Transfer Learning είναι τα AlexNet, VGG-16 και ResNet-152. Τα αποτελέσματα απέδωσαν ακρίβεια 85% με το ResNet και 82,5% ακρίβεια με το VGG-16, το οποίο υπερίσχυε του AlexNet κατά σχεδόν 10%. Η αντικατάσταση των πλήρως συνδεδεμένων στρωμάτων από έναν ταξινομητή Support Machine Vector δεν βελτίωσε την ακρίβεια ταξινόμησης. Τέλος τα επεξεργασμένα χαρακτηριστικά απέδωσαν πολύ χαμηλότερη ακρίβεια από τα βαθιά συνελκτικά δίκτυα.

- Keep Your AI-es on the Road: Tackling Distracted Driver Detection with Convolutional Neural Networks and Targeted Data Augmentation (Nikka Mofid, Jasmine Bayrooti, Shreya Ravi). Σε αυτήν την εργασία χρησιμοποίησαν προ-εκπαιδευμένα μοντέλα ταξινόμησης εικόνας, τεχνικές αύξησης δεδομένων, προ-επεξεργασία εικόνων μέσω της βιβλιοθήκης OpenCV και την προσέγγιση της τμηματοποίησης δέρματος (skin segmentation). Το καλύτερό τους μοντέλο περιλάμβανε όλα τα παραπάνω και πιο συγκεκριμένα, τμηματοποίηση δέρματος, θόλωση προσώπου και κλασικές τεχνικές αύξησης των δεδομένων. Αυτό το μοντέλο πέτυχε περίπου 15% αύξηση της βαθμολογίας F1 σε σχέση με ένα βασικό μοντέλο χωρίς τις παραπάνω τεχνικές, δείχνοντας έτσι την δυνατότητα που δίνουν αυτές οι τεχνικές για την ενίσχυση της ισχύος των νευρωνικών δικτύων.
- Classification of Driver Distraction (Samuel Colbran, Kaiqi Cen, Danni Luo). Σ' αυτό το άρθρο περιγράφεται τη διαδικασία χρήσης του πλαισίου βαθιάς εκμάθησης caffe για την εκπαίδευση και δοκιμή δύο μοντέλων νευρωνικών δικτύων VGG-16 και GoogleNet, για την ταξινόμηση των περισπασμένων οδηγών για την πρόκληση της State Farm στο Kaggle. Για τη βελτίωση των αποτελεσμάτων τους, δημιούργησαν ένα ensemble μοντέλων υπολογίζοντας τον μέσο όρο τους ενώ ύστερα εφάρμοσαν τον αλγόριθμο ταξινόμησης k-nearest neighbors. Το καλύτερο αποτέλεσμα είχε απώλεια 0,28554 χρησιμοποιώντας το συνδυασμένο σύνολο VGG-16 και GoogleNet με σταθμισμένος μέσο όρο $K = 10$. Αυτή η βαθμολογία τους έβαλε στο κορυφαίο 12% μεταξύ όλων των 1440 συμμετεχόντων στο Kaggle.
- Detection of Distracted Driver using Convolutional Neural Network (Bhakti Baheti, Suhas Gajre, Sanjay Talbar). Σε αυτή την εργασία τροποποίησαν την αρχιτεκτονική του VGG-16, ενώ εφάρμοσαν διάφορα τεχνικές κανονικοποίησης με σκοπό την βελτίωση της απόδοσης του μοντέλου. Πειραματικά αποτελέσματα έδειξαν ότι το σύστημά τους υπερετερούσε των προηγούμενων μεθόδων στη βιβλιογραφία επιτυγχάνοντας ακρίβεια 96,31% επεξεργάζοντας 42 εικόνες ανά δευτερόλεπτο σε GPU. Επίσης μελέτησαν την επίδραση του στρώματος Dropout, της κανονικοποίησης L2 και του στρώματος Batch Normalization στην απόδοση του συστήματος. Στη συνέχεια, παρουσίασαν μια τροποποιημένη έκδοση της αρχιτεκτονικής τους που

επιτυγχάνει 95,54% ακρίβεια ταξινόμησης με τον αριθμό των παραμέτρων μειωμένων από 140M του αρχικού VGG-16 σε μόνο 15M.

- Driver Distraction Identification with an Ensemble of Convolutional Neural Networks (Hesham M. Eraqi, Yehya Abouelnaga, Mohamed H. Saad, Mohamed N. Moustafa). Σε αυτό το άρθρο, παρουσιάστηκε το πρώτο δημόσια διαθέσιμο σύνολο δεδομένων για τον εντοπισμό της απόσπασης της προσοχής του οδηγού με περισσότερες στάσεις απόσπασης της προσοχής από τις υπάρχουσες εναλλακτικές λύσεις. Επιπλέον, πρότειναν μια αξιόπιστη λύση βασισμένη στη βαθιά μάθηση που επιτυγχάνει ακρίβεια 90%. Το σύστημα αποτελείται από ένα γενετικά σταθμισμένο σύνολο συνελκτικών νευρωνικών δικτύων και δείχνουν ότι ένα τέτοιο σύστημα που χρησιμοποιεί έναν γενετικό αλγόριθμο αποδίδει καλύτερα στην εργασία της ταξινόμησης. Ακόμη μελέτησαν την επίδραση διαφορετικών οπτικών στοιχείων στην ανίχνευση της απόσπασης της προσοχής μέσω εντοπισμών προσώπου, χεριών και τμηματοποίησης του δέρματος. Τέλος, παρουσίασαν μια έκδοση του συνόλου τους που μπορούσε να επιτύχει 84,64% ακρίβεια ταξινόμησης και να λειτουργεί σε περιβάλλον πραγματικού χρόνου.

2.5 Η μεθοδολογία

Σ' αυτήν την υποενότητα θα παρουσιαστεί ο διαδικασία επίλυσης του προβλήματος αφού πρώτα αποσαφηνιστούν κάποιες βασικές έννοιες που χρειάζονται για την πλήρη κατανόησή της.

2.5.1 Transfer Learning

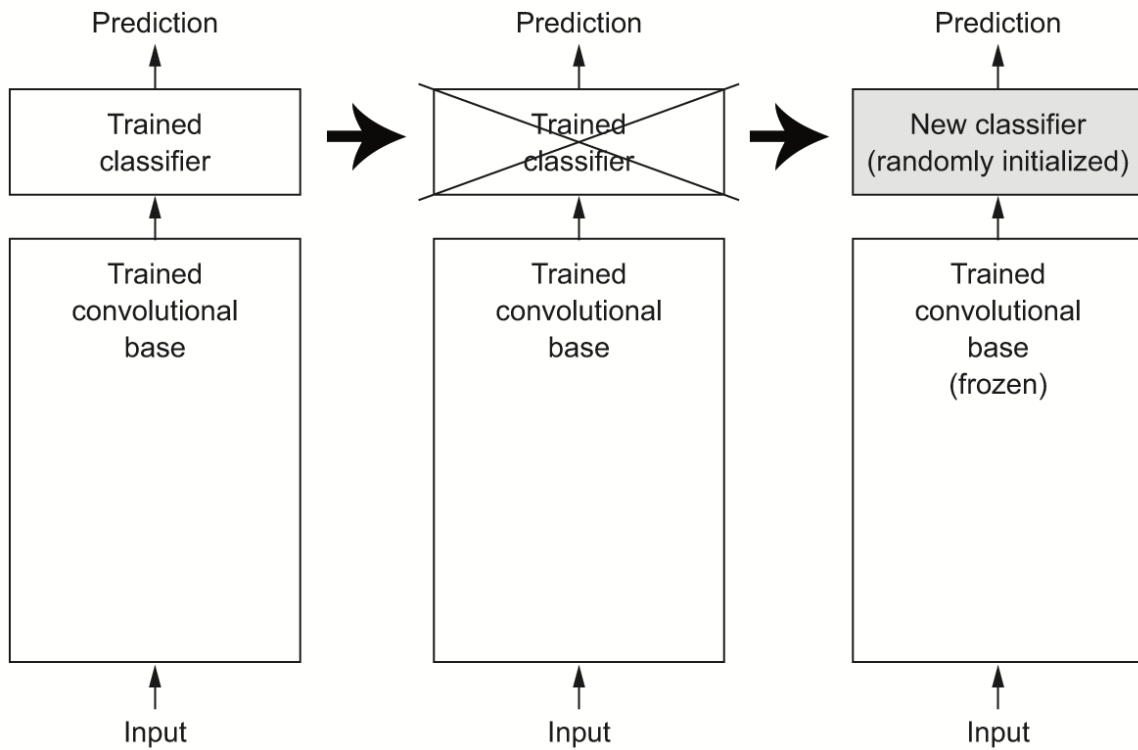
Τα νευρωνικά δίκτυα ενώ κάνουν πολύ καλά την δουλειά τους, έχουν ένα μεγάλο μειονέκτημα, χρειάζονται μεγάλο όγκο δεδομένων για να εκπαιδευτούν. Αυτήν η αδυναμία τους, καλύπτεται από το Transfer Learning. Transfer Learning είναι μια συχνή προσέγγιση για τέτοια ζητήματα, γιατί χρησιμοποιούνται προ-εκπαιδευμένα μοντέλα που έχουν εκπαιδευτεί σε ένα μεγάλο σύνολο δεδομένων και χρησιμοποιούνται είτε αυτούσια ως μοντέλα, είτε ως βάση για ένα νέο μοντέλο

Ο λογική του Transfer Learning είναι ότι όταν ένα μοντέλο έχει εκπαιδευτεί σε ένα αρκετά μεγάλο σύνολο δεδομένων, θα λειτουργήσει αποτελεσματικά και ως γενικό μοντέλο. Έτσι μπορούν να εκμεταλλευτούν οι ήδη υπάρχουσες γνώσεις του μοντέλου, χωρίς να χρειάζεται να εκπαιδευτεί ένα νέο μοντέλο από την αρχή σε ένα εξίσου μεγάλο σύνολο δεδομένων.

Υπάρχουν δύο τρόποι χειρισμού τέτοιων προ-εκπαιδευμένων μοντέλων:

1. Εξαγωγή Χαρακτηριστικών (Feature Extraction): Η εξαγωγή χαρακτηριστικών αποτελείται από τη χρήση των αναπαραστάσεων που έμαθε προηγουμένως ένα άλλο εκπαιδευμένο μοντέλο για να εξαχθούν ενδιαφέροντα χαρακτηριστικά από νέα δεδομένα.

Αυτά τα χαρακτηριστικά στη συνέχεια περνούν μέσα από έναν νέο ταξινομητή, ο οποίος εκπαιδεύεται από την αρχή. Δεν χρειάζεται να εκπαιδευτεί ξανά ολόκληρο το μοντέλο.



Εικόνα 37: Ανταλλαγή ταξινομητών διατηρώντας την ίδια συνεκτική βάση. [13]

Το βασικό συνεκτικό δίκτυο περιέχει ήδη χαρακτηριστικά που είναι γενικά χρήσιμα για την ταξινόμηση εικόνων. Επίσης ο λόγος που επαναχρησιμοποιείται μόνο η συνεκτική βάση και όχι ο ταξινομητής του προ-εκπαιδευμένου μοντέλου, είναι ότι οι αναπαραστάσεις που έχουν μαθευτεί από τη συνεκτική βάση είναι πιθανό να είναι πιο γενικές και επομένως πιο επαναχρησιμοποιήσιμες. Αντίθετα οι αναπαραστάσεις που μαθαίνει ο ταξινομητής είναι απαραίτητα συγκεκριμένες για το σύνολο των κατηγοριών στις οποίες εκπαιδεύτηκε το μοντέλο.

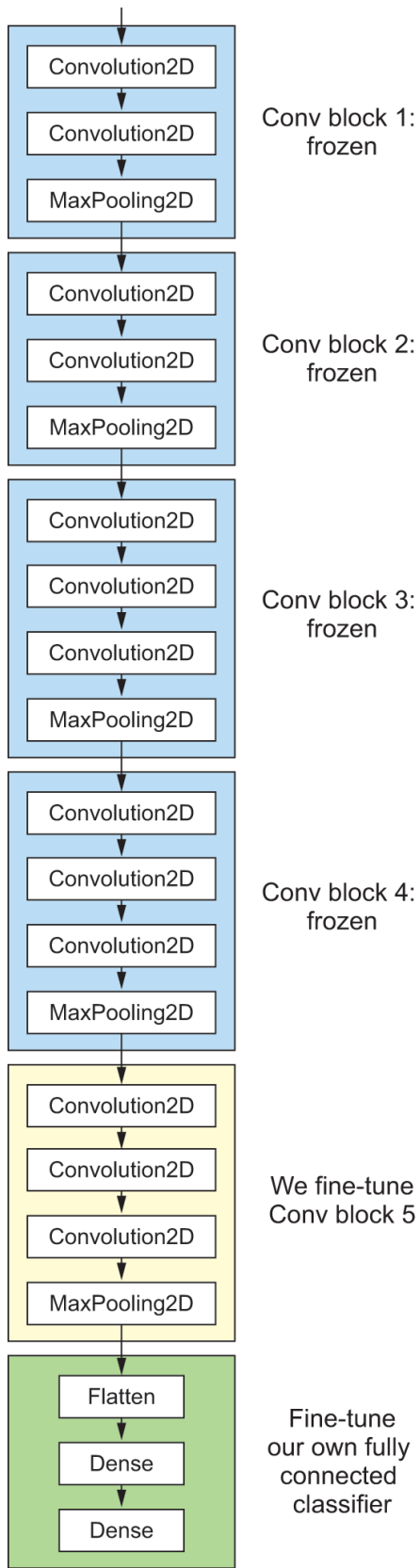
Επιπλέον, οι αναπαραστάσεις που βρίσκονται σε πυκνά συνδεδεμένα στρώματα δεν περιέχουν πλέον καμία πληροφορία σχετικά με το πού βρίσκονται τα αντικείμενα στην εικόνα εισόδου.

Τέλος, όπως αναφέρθηκε και στην υπό-ενότητα 1.3.7 του 1^{ου} Κεφαλαίου οι αναπαραστάσεις που εξάγουν τα πρώτα στρώματα είναι πιο γενικές (γραμμές, καμπύλες, γωνίες), ενώ τα στρώματα που βρίσκονται ψηλότερα εξάγουν πιο αφηρημένες έννοιες (όπως μάτι, στόμα, αυτί).

Έτσι, εάν το νέο σύνολο δεδομένων διαφέρει πολύ από το σύνολο δεδομένων στο οποίο

εκπαιδύτηκε το αρχικό μοντέλο, ίσως είναι προτιμότερο να χρησιμοποιηθούν μόνο τα πρώτα στρώματα του μοντέλου για να γίνει η εξαγωγή χαρακτηριστικών, αντί να χρησιμοποιηθεί ολόκληρη τη βάση μετατροπής.

2. Fine-Tuning: Το fine-tuning συνιστάται από το «ξεπάγωμα» μερικών από τα ανώτερα στρώματα μια παγωμένης βάσης μοντέλου που χρησιμοποιείται για την εξαγωγή χαρακτηριστικών, και εκπαιδύεται μαζί με το νέο τμήμα του μοντέλου που προστέθηκε. Λέγεται έτσι γιατί προσαρμόζει τις ελαφρώς πιο αφηρημένες αναπαραστάσεις του προ-εκπαιδευμένου μοντέλου προκειμένου να γίνουν πιο σχετικές με το νέο πρόβλημα. [13]



Εικόνα 38:Βελτιστοποίηση του τελευταίου συνελκτικού μπλοκ του δικτύου VGG16. [13]

Είναι πολύ σημαντικό να παγώσει η συνελικτική βάση για να μπορέσει ξεκινήσει η εκπαίδευση του ταξινομητή με τα τυχαία βάρη στην κορυφή του δικτύου. Εάν ο ταξινομητής δεν έχει ήδη εκπαιδευτεί, τότε το σήμα σφάλματος που διαδίδεται στο δίκτυο κατά τη διάρκεια της εκπαίδευσης θα είναι πολύ μεγάλο και οι αναπαραστάσεις που έχουν μαθευτεί προηγουμένως από τα επίπεδα θα χαθούν. Κατά συνέπεια τα βήματα για τη σωστή ρύθμιση ενός δικτύου είναι τα εξής:

- I. Εισαγωγή του προσαρμοσμένου δικτύου πάνω στο προ-εκπαιδευμένο δίκτυο.
- II. Πάγωμα του προ-εκπαιδευμένου δικτύου βάσης.
- III. Εκπαίδευση του προσαρμοσμένου δικτύου που προστέθηκε.
- IV. Ξεπάγωμα ορισμένων επιπέδων του προ-εκπαιδευμένου δικτύου βάσης (εκτός από τα επίπεδα “batch normalization”).
- V. Εκπαίδευση από κοινού των ξεπαγωμένων επιπέδων και αυτών που προστέθηκαν.

Εδώ γεννάται και η εύλογη ερώτηση «γιατί να μην ξεπαγώσει ολόκληρο το προ-εκπαιδευμένο δίκτυο βάσης μετά την πρώτη εκπαίδευση;». Γίνεται αυτό αλλά, πρέπει να ληφθούν υπόψιν τα ακόλουθα:

- Όπως έχει αναφερθεί πολλάκις τα κάτω στρώματα της συνελικτικής βάσης κωδικοποιούν γενικά και επαναληπτικά χαρακτηριστικά, ενώ τα στρώματα που βρίσκονται ψηλά κωδικοποιούν πιο εξειδικευμένα χαρακτηριστικά. Επομένως έχει πιο νόημα να ρυθμιστούν μόνο τα ψηλά στρώματα για επαναπροσδιοριστούν στο υπάρχον πρόβλημα. Ακόμη είναι πιθανό να υπάρχουν επιπτώσεις στις αποδόσεις του δικτύου αν ρυθμιστούν τα χαμηλά επίπεδα.
- Όσο περισσότερες παράμετροι εκπαιδεύονται, τόσο περισσότερο αυξάνεται κίνδυνος εμφάνισης της υπερβολικής προσαρμογής (overfitting).

Γι’ αυτό είναι μια καλή στρατηγική να ξεπαγώνουν μόνο τα δύο ή τρία πρώτα στρώματα της συνελικτικής βάσης. [13]

Από τους δύο τρόπους προτιμήθηκε ο δεύτερος γιατί, αν και πιο απαιτητικός φέρνει καλύτερα αποτελέσματα από τον πρώτο.

2.5.2 Μετρήσεις Αξιολόγησης

Για την σωστή αξιολόγηση του κάθε μοντέλου χρειάζεται να ποσοτικοποιηθεί η ποιότητα των προβλέψεων. Αυτό σημαίνει ότι υπάρχουν πολλοί διαφορετικοί τρόποι μέτρησης της απόδοσης ενός μοντέλου, που όλοι μαζί δίνουν μια πιο πολύπλευρη εικόνα της συμπεριφοράς τους.

Συγκεκριμένα για προβλήματα κατηγοριοποίησης αυτοί που χρησιμοποιούνται πιο συχνά είναι:

- Precision Score (Βαθμολογία Ακρίβειας): Η ακρίβεια είναι ο λόγος $\frac{tp}{(tp+fp)}$ όπου tp είναι ο αριθμός των πραγματικών θετικών (true positives) και fp ο αριθμός των ψευδώς θετικών (false positives). Η ακρίβεια είναι διαισθητικά η ικανότητα του ταξινομητή να μην χαρακτηρίζει ως θετικό ένα αρνητικό δείγμα. Η καλύτερη τιμή είναι 1 και η χειρότερη τιμή είναι 0.
- Recall Score (Βαθμολογία Ανάκλησης): Η ανάκληση είναι ο λόγος $\frac{tp}{(tp+fn)}$ όπου tp είναι ο αριθμός των πραγματικών θετικών (true positives) και fn ο αριθμός των ψευδώς αρνητικών (false negatives). Η ανάκληση είναι διαισθητικά η ικανότητα του ταξινομητή να βρει όλα τα θετικά δείγματα. Η καλύτερη τιμή είναι 1 και η χειρότερη τιμή είναι 0.
- F1 Score (Βαθμολογία F1): Η βαθμολογία F1 μπορεί να ερμηνευθεί ως σταθμισμένος μέσος όρος (weighted average) της ακρίβειας και της ανάκλησης, όπου φτάνει την καλύτερη της τιμή στο 1 και τη χειρότερη στο 0. Η σχετική συμβολή της ακρίβειας και της ανάκλησης στη βαθμολογία F1 είναι ίση. Ο τύπος για τη βαθμολογία F1 είναι:
$$F1 = 2 * \frac{(precision * recall)}{(precision + recall)}$$
- Accuracy Score (Βαθμολογία Ακρίβειας): Βαθμολογία ακρίβειας ταξινόμησης.

		Predicted Label	
		Positive	Negative
Actual Label	Positive	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
	Negative	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

Εικόνα 39: Παράδειγμα πραγματικών και μη, θετικών και αρνητικών

2.5.3 Προ-εκπαιδευμένο μοντέλο VGG16

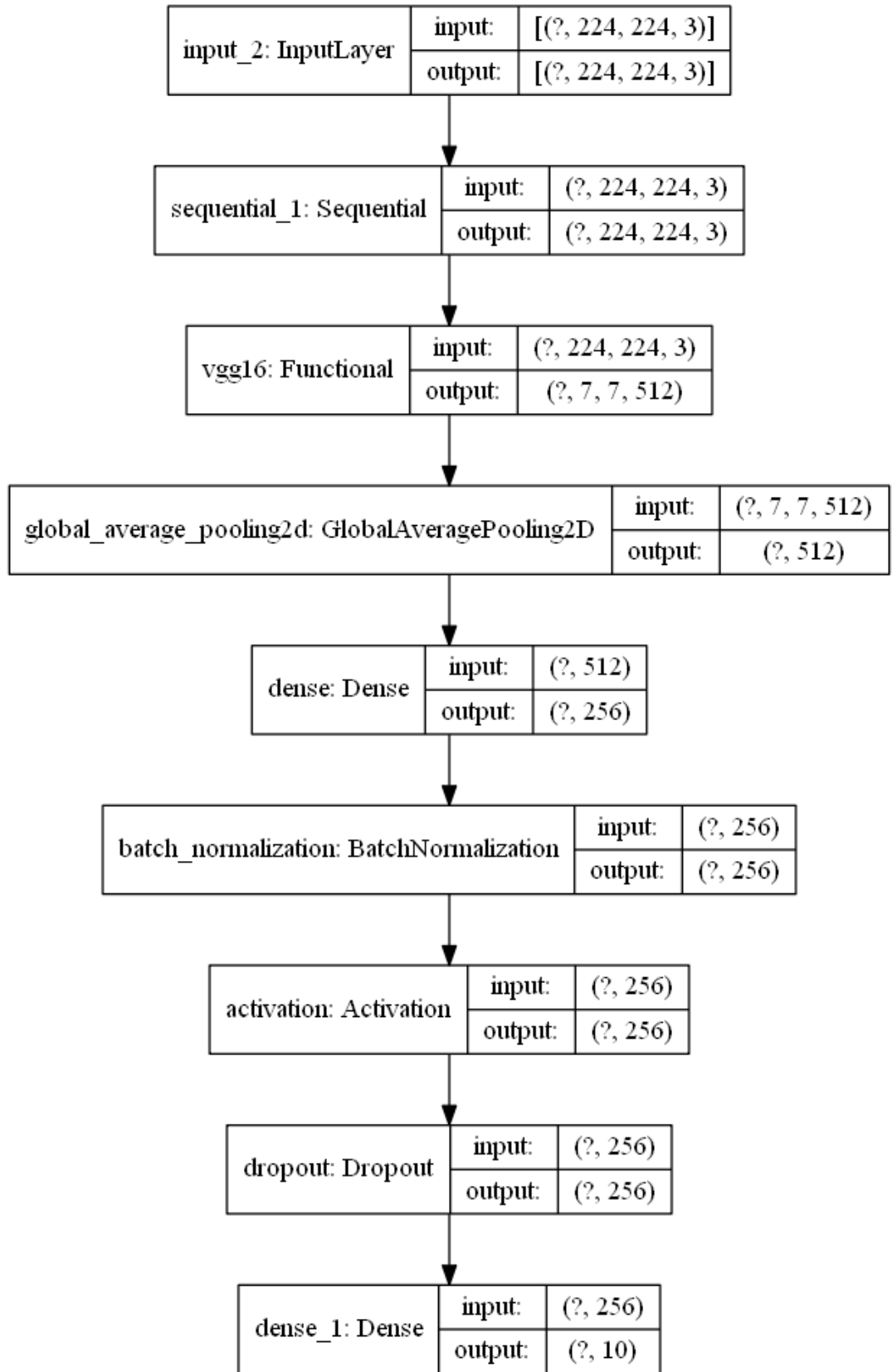
Το προ-εκπαιδευμένο μοντέλο VGG16 χρησιμοποιήθηκε ως βάση ενός νέου μοντέλου. Το πάνω στρώμα δεν συμπεριλήφθηκε γιατί προστέθηκε άλλη έξοδος. Κρατήθηκαν όμως τα βάρη τα οποία είχε από την εκπαίδευσή του πάνω στο σύνολο δεδομένων Imagenet. Επίσης όλα τα στρώματα του μοντέλου πάγωσαν ώστε να μην αλλάξουν τα βάρη κατά την πρώτη εκπαίδευση. Μετά το 5^ο κομμάτι στρωμάτων προστέθηκαν με την σειρά τα εξής στρώματα: Global Average Pooling, Dense με 256 μονάδες, Batch Normalization, ReLu Activation, Dropout με ρυθμό 0.5 και τέλος η έξοδος που είναι ένα στρώμα Dense με 10 μονάδες όσες και οι κατηγορίες και με Activation Softmax.

```
Model: "vgg16"
```

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0

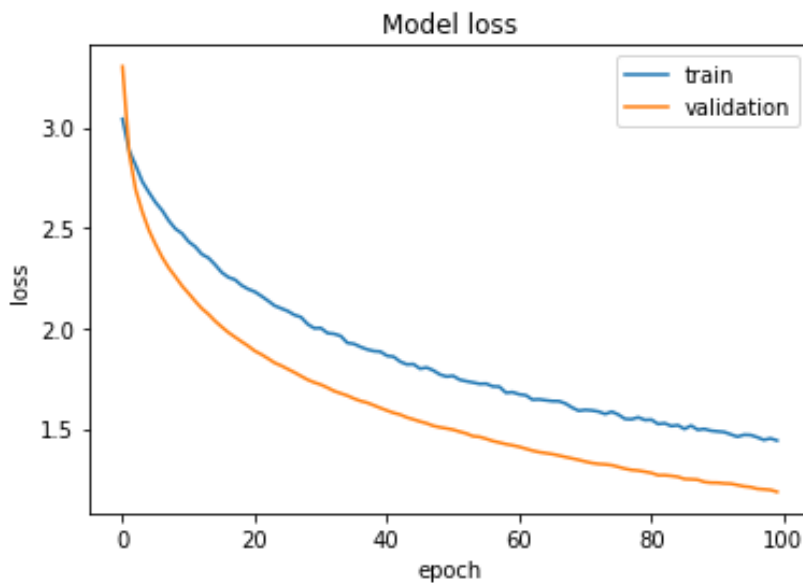
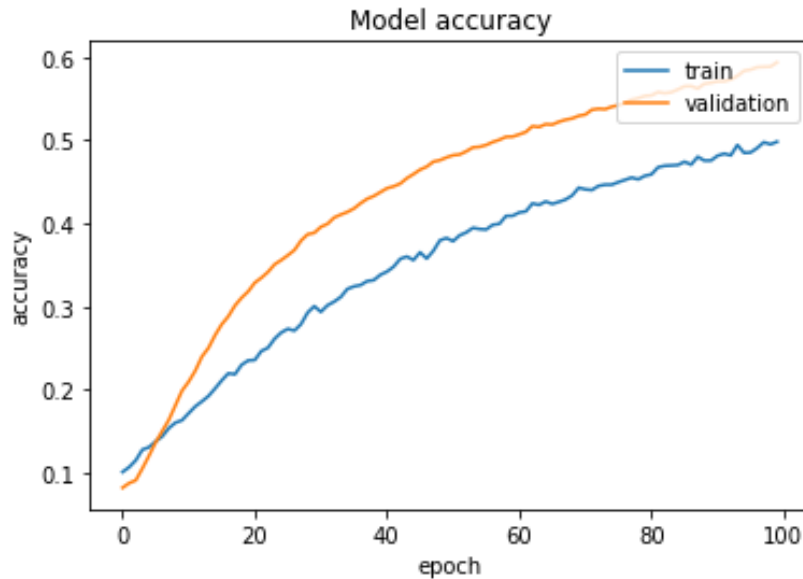
```
Total params: 14,714,688  
Trainable params: 0  
Non-trainable params: 14,714,688
```

Εικόνα 40: Η Vgg16 βάση για το μοντέλο.



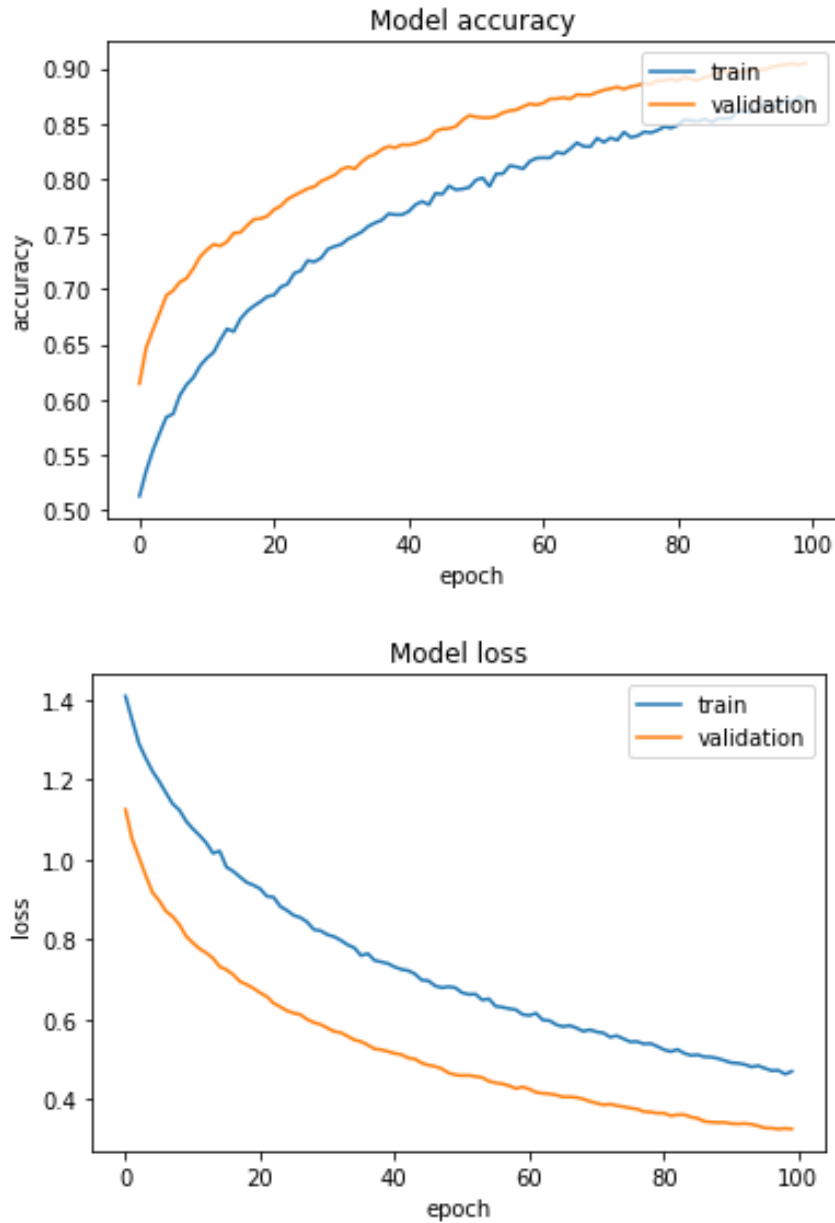
Εικόνα 41: Η αρχιτεκτονική του τελικού μοντέλου

Ο βελτιστοποιητής (Optimizer) που επιλέχθηκε είναι ο SGD (Gradient Descent optimizer) με learning rate 10^{-4} , momentum 0.9 και batch size 128. Το μοντέλο εκπαιδεύτηκε με παγωμένα τα στρώματα της βάσης για 100 εποχές και έφτασε στα εξής αποτελέσματα: loss: 1.4421 - accuracy: 0.4984 - val_loss: 1.1868 - val_accuracy: 0.5938. Επίσης η εκπαίδευση έγινε πάνω στο πρωτότυπο dataset.



Εικόνα 42: Η γραφική παράσταση ακρίβειας και απώλειας του μοντέλου VGG16.

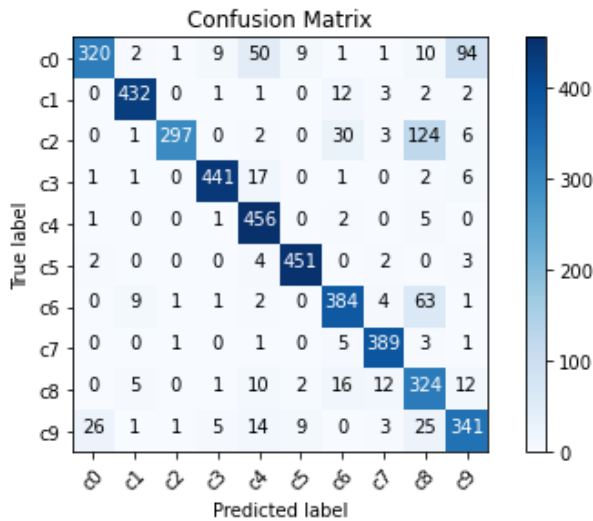
Έπειτα ξεπάγωσε το 5^ο κομμάτι στρωμάτων ενώ το learning rate μειώθηκε στην τιμή 10^{-5} χρησιμοποιώντας τον ίδιο optimizer. Η εκπαίδευση συνέχισε για άλλες 100 εποχές. Συνολικά στην 200η εποχή το μοντέλο έφτασε στα εξής αποτελέσματα: loss: 0.4698 - accuracy: 0.8717 - val_loss: 0.3259 - val_accuracy: 0.9047



Εικόνα 43: Η γραφική παράσταση ακρίβειας και απώλειας του μοντέλου VGG16 στην 2η εκπαίδευση.

Ακόμη το μοντέλο αξιολογήθηκε πάνω στο validation set με τα ακόλουθα αποτελέσματα: loss: 0.4460 - accuracy: 0.8560 → Test accuracy: 0.856

Το Confusion Matrix μας δίνει μια εκτενέστερη εικόνα της απόδοσης του μοντέλου.



Εικόνα 44: Το confusion matrix του μοντέλου VGG16.

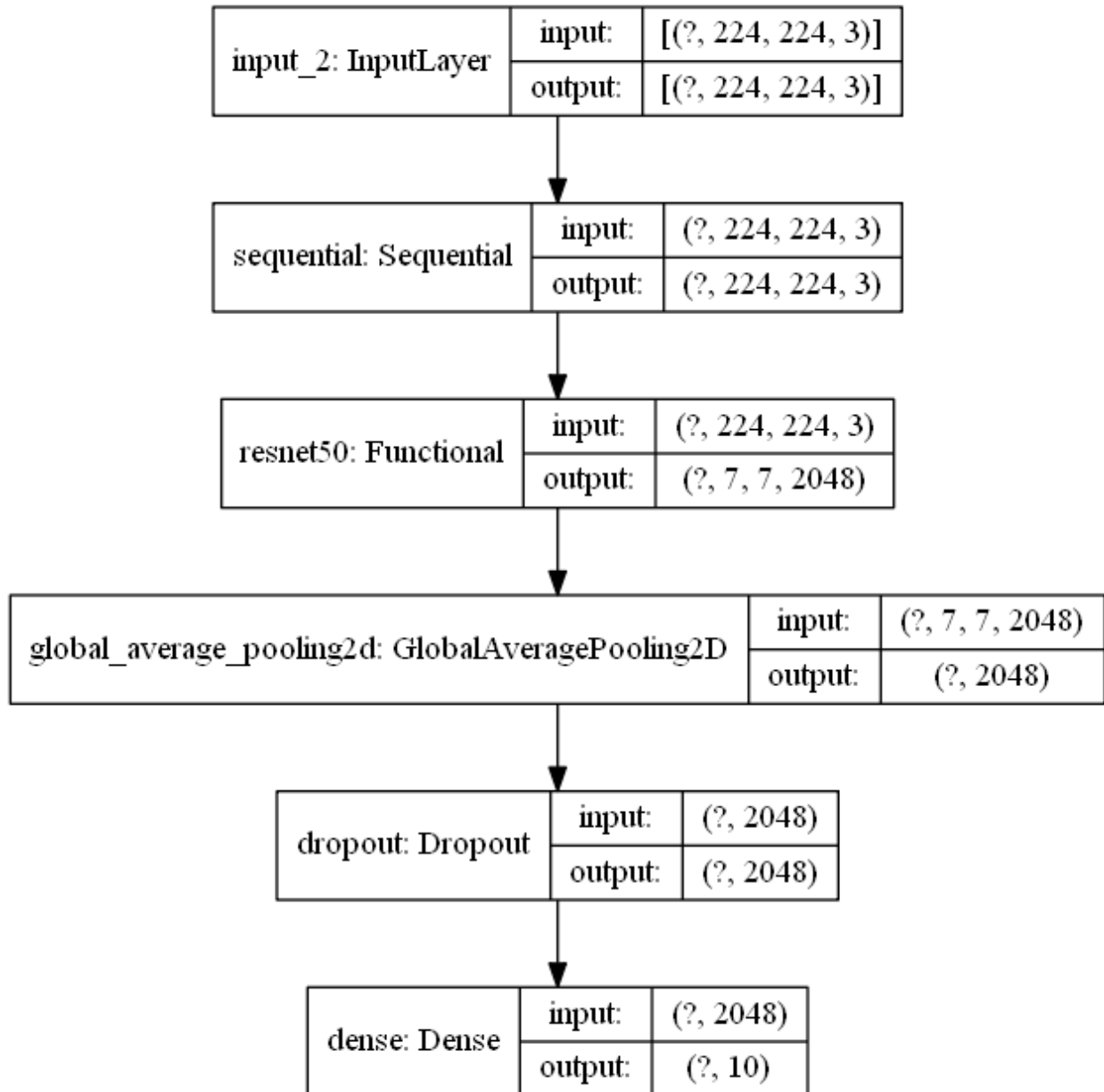
Στο παραπάνω πίνακα φαίνεται πόσες φορές το μοντέλο έκανε σωστή πρόβλεψη και πόσες φορές μπερδέψε μία κατηγορία με μία άλλη. Διακρίνεται εύκολα, ότι το μοντέλο ταξινόμησε λάθος την κατηγορία c0 με την c3, c4 και c5, 11, 43 και 16 φορές αντίστοιχα. Την c1 με τις c6 και c8, 31 και 11 φορές. Την c3 με την c4, 23 φορές. Την c6 με την c8, 61 φορές. Την c7 με την c8, 16 φορές. Την c8 με τις c2, c4 και c8, 15, 13, 18 φορές αντίστοιχα και την c9 με τις c0, c5, c7 και c8, 37, 10, 15, 31 φορές αντίστοιχα. Οι περιπτώσεις που το μοντέλο ταξινόμησε λάθος μια κατηγορία λιγότερο από 10 φορές, κρίνονται αμελητέες.

	precision	recall	f1-score	support
0	0.91	0.64	0.76	497
1	0.96	0.95	0.96	453
2	0.99	0.64	0.78	463
3	0.96	0.94	0.95	469
4	0.82	0.98	0.89	465
5	0.96	0.98	0.97	462
6	0.85	0.83	0.84	465
7	0.93	0.97	0.95	400
8	0.58	0.85	0.69	382
9	0.73	0.80	0.77	425
accuracy			0.86	4481
macro avg	0.87	0.86	0.85	4481
weighted avg	0.87	0.86	0.86	4481
Precision score is:	0.87			
Recall score is:		0.86		
f1 score score is:			0.86	
Accuracy score is:				0.86

Εικόνα 45: Πίνακας αξιολόγησης του μοντέλου VGG16.

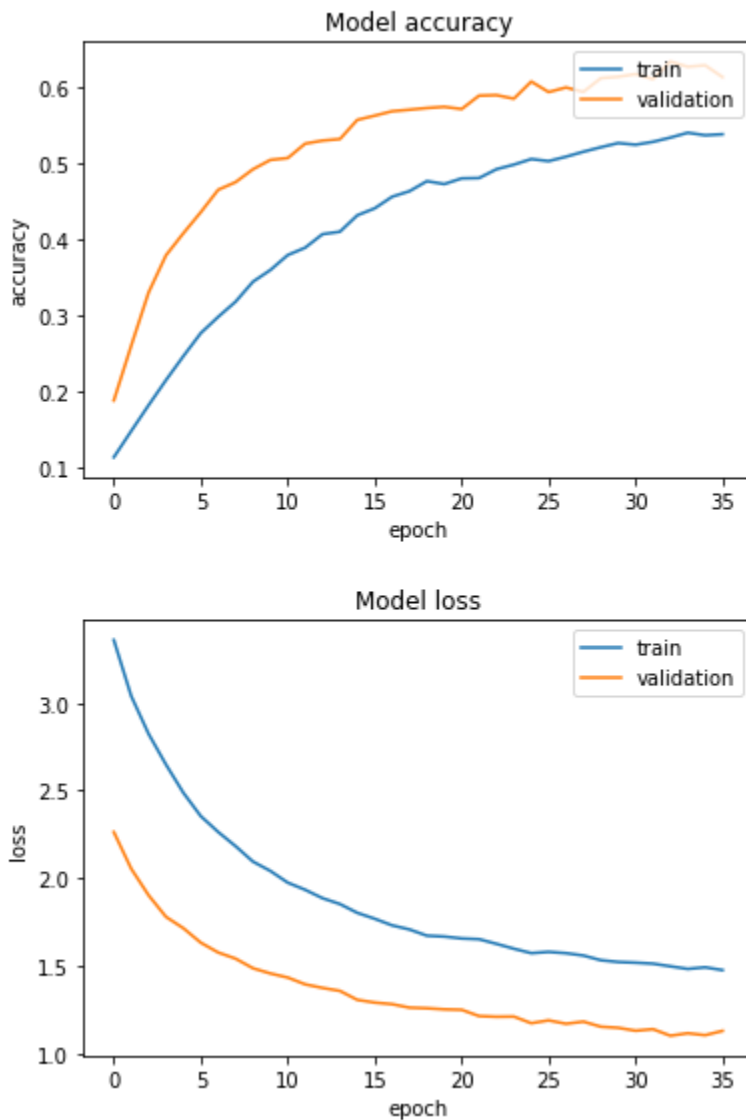
2.5.4 Προ-εκπαιδευμένο μοντέλο ResNet50

Παρόμοια διαδικασία ακολουθήθηκε για την ανάπτυξη μοντέλου έχοντας ως βάση το μοντέλο ResNet50. Η μόνη διαφορά με πριν είναι ότι μετά το τελευταίο κομμάτι στρωμάτων προστέθηκαν με την σειρά τα εξής στρώματα: Global Average Pooling 2D, Dropout με ρυθμό 0.5 και Dense με 10 μονάδες όσες και οι έξοδοι.



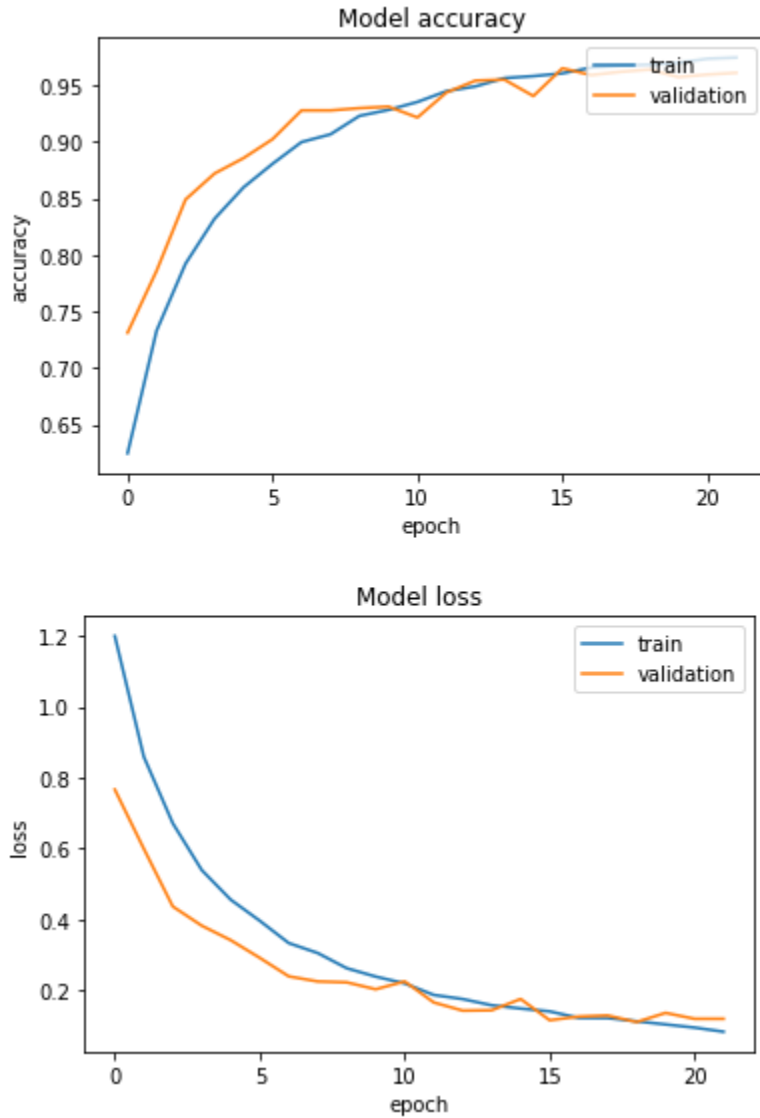
Εικόνα 46: Η αρχιτεκτονική του 2^{ου} τελικού μοντέλου

Ο βελτιστοποιητής (Optimizer) που επιλέχθηκε είναι ο Adam με learning rate 10^{-4} και batch size 128. Το μοντέλο αρχικά εκπαιδεύτηκε για 36 εποχές με παγωμένα τα στρώματα της βάσης. Τα αποτελέσματα είναι τα εξής: loss: 1.4752 - accuracy: 0.5382 - val_loss: 1.1283 - val_accuracy: 0.6134. Να σημειωθεί ότι η εκπαίδευση έγινε πάνω στο πρωτότυπο dataset με θολωμένα τα πρόσωπα των οδηγών.



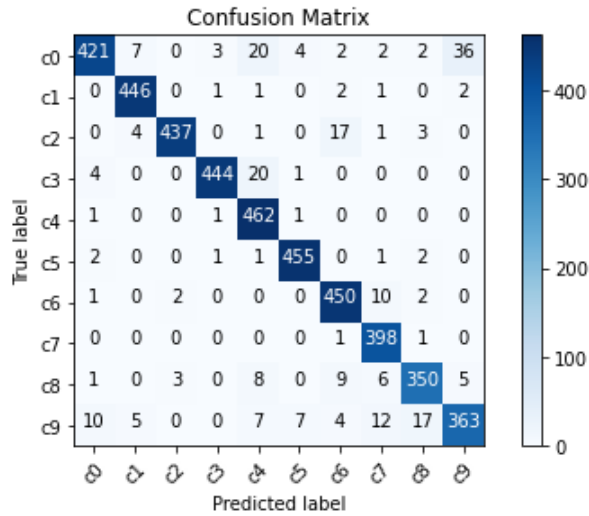
Εικόνα 47: Η γραφική παράσταση ακρίβειας και απώλειας του μοντέλου ResNet50.

Έπειτα ξεπάγωσαν 32 στρώματα ξεκινώντας από την κορυφή της βάσης resnet50, κρατώντας όμως όλα τα Batch Normalization στρώματα παγωμένα. Το μοντέλο συνέχισε την εκπαίδευσή του για άλλες 22 εποχές με χαμηλότερο learning rate με τιμή 10^{-5} . Τα αποτελέσματα είναι τα εξής: loss: 0.0834 - accuracy: 0.9747 - val_loss: 0.1199 - val_accuracy: 0.9612.



Εικόνα 48: Η γραφική παράσταση ακρίβειας και απώλειας του μοντέλου ResNet50 στην 2η εκπαίδευση.

Η αξιολόγηση του μοντέλου έγινε πάνω στο validation set και έφερε τα ακόλουθα αποτελέσματα: loss: 0.1108 - accuracy: 0.9641 → Test accuracy: 0.964



Εικόνα 49: Το confusion matrix του μοντέλου ResNet50.

Με μία γρήγορη ματιά, φαίνεται ότι τα αποτελέσματα είναι καλύτερα από αυτά του προηγούμενου μοντέλου. Πιο αναλυτικά, το μοντέλο ταξινόμησε λάθος την κατηγορία c0 με την c9, 10 φορές. Την c4 με τις c3 και c0, 20 φορές έκαστοι. Την c6 με την c2, 17 φορές. Την c7 με την c6 και c9, 10 και 12 φορές αντίστοιχα. Την c8 με την c9, 17 φορές. Τέλος την c9 με την c1, 36 φορές. Οι περιπτώσεις που το μοντέλο ταξινόμησε λάθος μια κατηγορία λιγότερο από 10 φορές, κρίνονται αμελητέες.

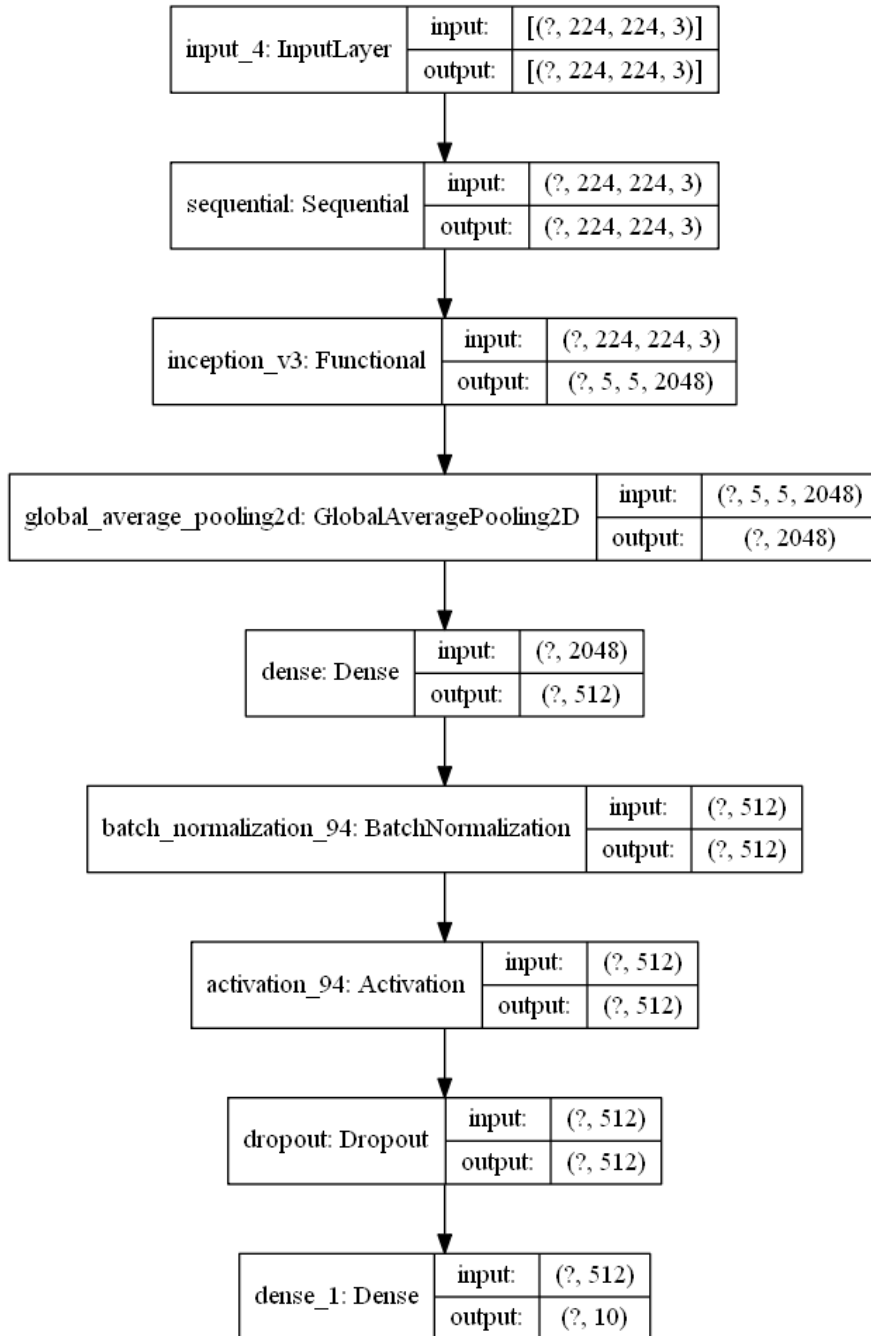
	precision	recall	f1-score	support
0	0.96	0.85	0.90	497
1	0.97	0.98	0.97	453
2	0.99	0.94	0.97	463
3	0.99	0.95	0.97	469
4	0.89	0.99	0.94	465
5	0.97	0.98	0.98	462
6	0.93	0.97	0.95	465
7	0.92	0.99	0.96	400
8	0.93	0.92	0.92	382
9	0.89	0.85	0.87	425
accuracy			0.94	4481
macro avg	0.94	0.94	0.94	4481
weighted avg	0.94	0.94	0.94	4481

Precision score is: 0.94
 Recall score is: 0.94
 f1 score score is: 0.94
 Accuracy score is: 0.94

Εικόνα 50: Πίνακας αξιολόγησης του μοντέλου ResNet50

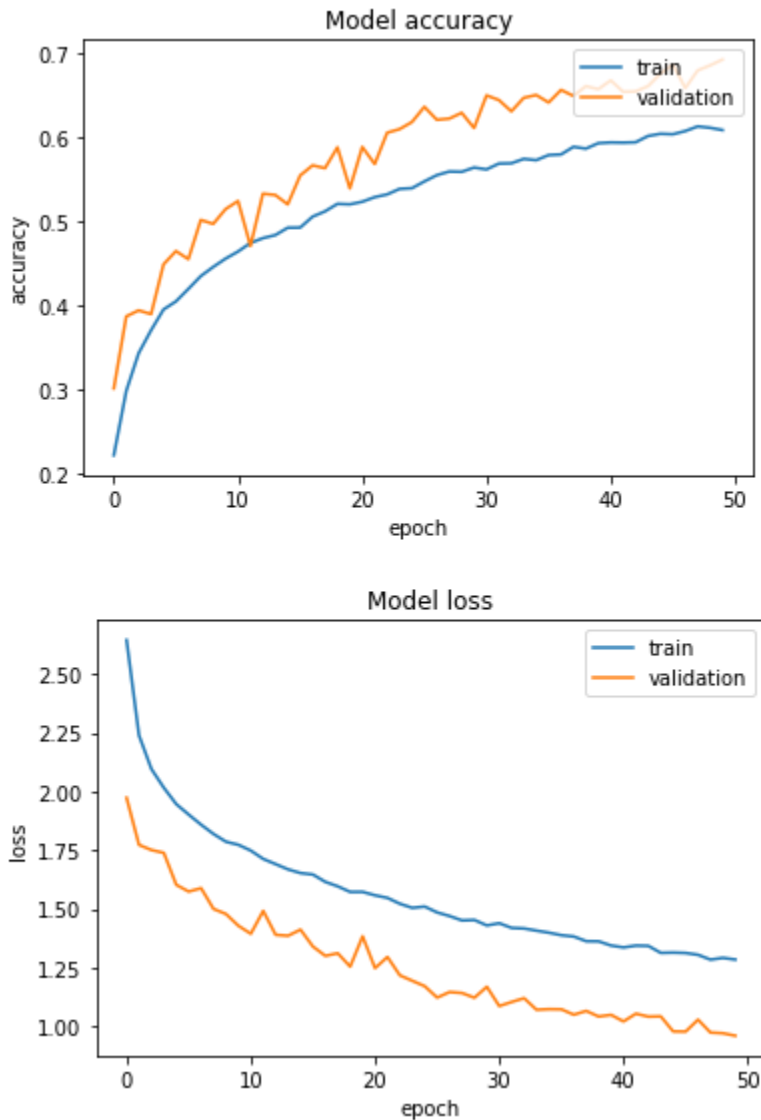
2.5.5 Προ-εκπαιδευμένο μοντέλο InceptionV3

Παρόμοια διαδικασία ακολουθήθηκε για την ανάπτυξη μοντέλου έχοντας ως βάση το μοντέλο InceptionV3. Η μόνη διαφορά με πριν είναι ότι μετά το τελευταίο κομμάτι στρωμάτων προστέθηκαν με την σειρά τα εξής στρώματα: Global Average Pooling 2D, Dense με 512 μονάδες, Batch Normalization, Activation ReLu, Dropout με ρυθμό 0.5 και Dense με 10 μονάδες όσες και οι έξοδοι.



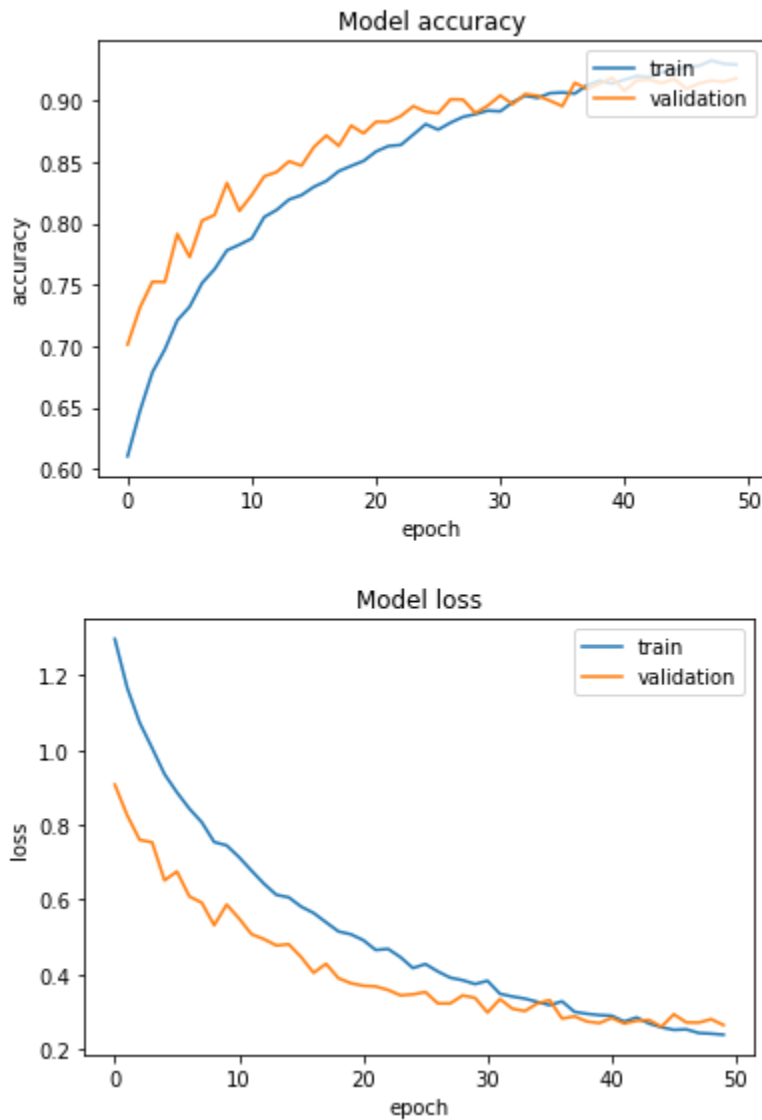
Εικόνα 51: Η αρχιτεκτονική του μοντέλου με βάση το InceptionV3

Ο βελτιστοποιητής (Optimizer) που επιλέχθηκε είναι ο Adam με learning rate 10^{-3} και batch size 128. Το μοντέλο αρχικά εκπαιδεύτηκε για 50 εποχές με παγωμένα τα στρώματα της βάσης. Τα αποτελέσματα είναι τα εξής: loss: 1.2838 - accuracy: 0.6092 - val_loss: 0.9591 - val_accuracy: 0.6931. Η εκπαίδευση έγινε πάνω στο τμηματοποιημένο dataset.



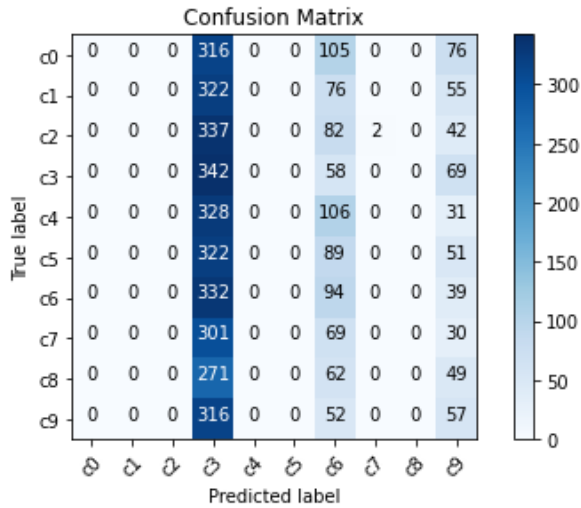
Εικόνα 52: Η γραφική παράσταση ακρίβειας και απώλειας του μοντέλου InceptionV3.

Ύστερα ξεπάγωσαν 31 στρώματα ξεκινώντας από την κορυφή της βάσης inceptionv3, κρατώντας όμως όλα τα Batch Normalization στρώματα παγωμένα. Το μοντέλο συνέχισε την εκπαίδευσή του για άλλες 50 εποχές με χαμηλότερο learning rate με τιμή 10^{-4} . Τα αποτελέσματα είναι τα εξής: loss: 0.2380 - accuracy: 0.9290 - val_loss: 0.2638 - val_accuracy: 0.9179.



Εικόνα 53: Η γραφική παράσταση ακρίβειας και απώλειας του μοντέλου InceptionV3 στην 2η εκπαίδευση.

Η αξιολόγηση του μοντέλου έγινε πάνω στο validation set και έφερε τα ακόλουθα αποτελέσματα: loss: 15.2563 - accuracy: 0.1100 → Test accuracy: 0.1100



Εικόνα 54: Το confusion matrix του μοντέλου InceptionV3.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	497
1	0.00	0.00	0.00	453
2	0.00	0.00	0.00	463
3	0.11	0.73	0.19	469
4	0.00	0.00	0.00	465
5	0.00	0.00	0.00	462
6	0.12	0.20	0.15	465
7	0.00	0.00	0.00	400
8	0.00	0.00	0.00	382
9	0.11	0.13	0.12	425
accuracy			0.11	4481
macro avg	0.03	0.11	0.05	4481
weighted avg	0.03	0.11	0.05	4481

Precision score is: 0.03
 Recall score is: 0.11
 f1 score score is: 0.05
 accuracy score is: 0.11

Εικόνα 55: Πίνακας αξιολόγησης του μοντέλου InceptionV3

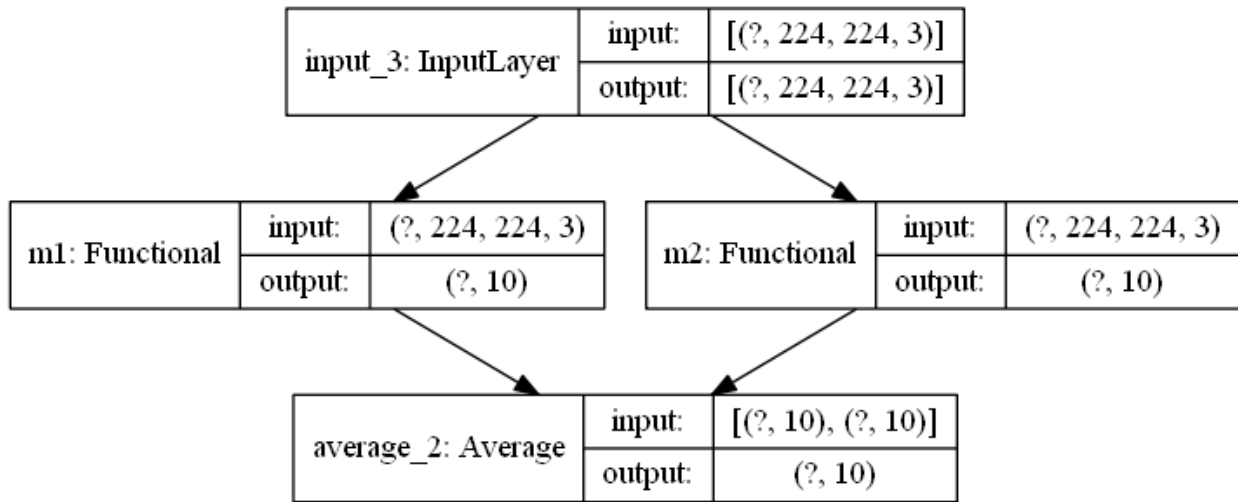
Παρατηρώντας τους παραπάνω πίνακες, γίνεται σαφές ότι το μοντέλο πάσχει από σοβαρό underfitting, δηλαδή δεν έχει μάθει τίποτα. Αυτό πιθανώς οφείλεται στο γεγονός ότι το μοντέλο τύπου U-net που εκπαιδεύτηκε για να πραγματοποιήσει το body segmentation στα δεδομένα μας, δεν έκανε καλά τη δουλειά του.

2.5.6 Model Ensemble

Στην στατιστική καθώς και στην μηχανική μάθηση οι μέθοδοι του ensemble χρησιμοποιούν πολλαπλούς αλγορίθμους μάθησης, για να πετύχουν καλύτερη προγνωστική απόδοση από αυτήν που θα μπορούσε να ληφθεί από οποιονδήποτε από τους αλγόριθμους μάθησης ξεχωριστά. [21]

Δυστυχώς δεν είναι δυνατή η πραγματοποίηση του ensemble όλων των αλγορίθμων μαζί, διότι η κάρτα γραφικών δεν έχει τη απαραίτητη ποσότητα μνήμης ώστε να μπορέσουν να υπολογιστούν τα διάφορα μετρικά όπως το confusion matrix. Κατά συνέπεια, επιλέχθηκαν μόνο τα μοντέλα VGG16 και ResNet50. Επίσης, το μοντέλο InceptionV3, είχε κακές επιδόσεις που δεν θα επιδρούσαν θετικά στο ensemble, άρα δεν θα χρησιμοποιούνταν ούτως ή άλλως.

Η διαδικασία που εκτελεί η μέθοδος ensemble, είναι να συλλέγει τις εξόδους των μοντέλων σε μια λίστα και ύστερα να βρίσκει την μέση τιμή τους.



Εικόνα 56: Η αρχιτεκτονική του ensemble των μοντέλων.

Model: "ensemble"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, 224, 224, 3)	0	
m1 (Functional) [0][0]	(None, 10)	14849610	input_1
m2 (Functional) [0][0]	(None, 10)	23608202	input_1
average (Average)	(None, 10)	0	m1[0][0] m2[0][0]

=====
=====
Total params: 38,457,812
Trainable params: 38,404,180
Non-trainable params: 53,632
=====

Εικόνα 57: Μια σύνοψη του Ensemble των δύο μοντέλων.

Οι επιδόσεις του μοντέλου δοκιμάστηκαν στο αρχικό dataset, αλλά και στο επεξεργασμένο. Τα αποτελέσματα αναγράφονται παρακάτω.

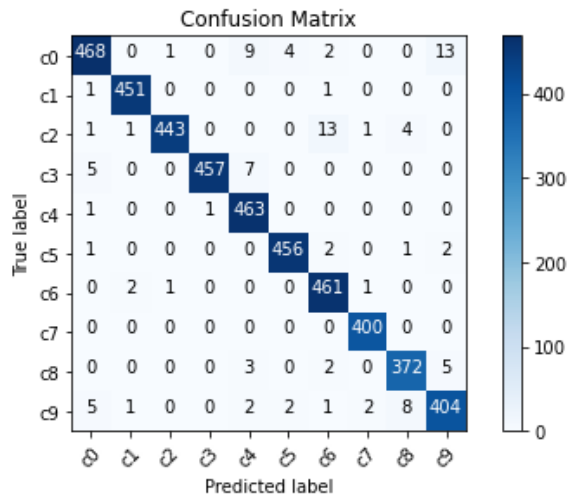
2.5.6.1 Αξιολόγηση του Ensemble των μοντέλου

Η αξιολόγηση του μοντέλου έγινε πάνω στο validation set και έφερε τα ακόλουθα αποτελέσματα: loss: 0.1980 - accuracy: 0.9763 → Test accuracy: 0.976

	precision	recall	f1-score	support
0	0.97	0.94	0.96	497
1	0.99	1.00	0.99	453
2	1.00	0.96	0.98	463
3	1.00	0.97	0.99	469
4	0.96	1.00	0.98	465
5	0.99	0.99	0.99	462
6	0.96	0.99	0.97	465
7	0.99	1.00	1.00	400
8	0.97	0.97	0.97	382
9	0.95	0.95	0.95	425
accuracy			0.98	4481
macro avg	0.98	0.98	0.98	4481
weighted avg	0.98	0.98	0.98	4481

Precision score is: 0.98
 Recall score is: 0.98
 f1 score is: 0.98
 Accuracy score is: 0.98

Εικόνα 58: Πίνακας αξιολόγησης του μοντέλου ensemble στο επεξεργασμένο σύνολο δεδομένων.

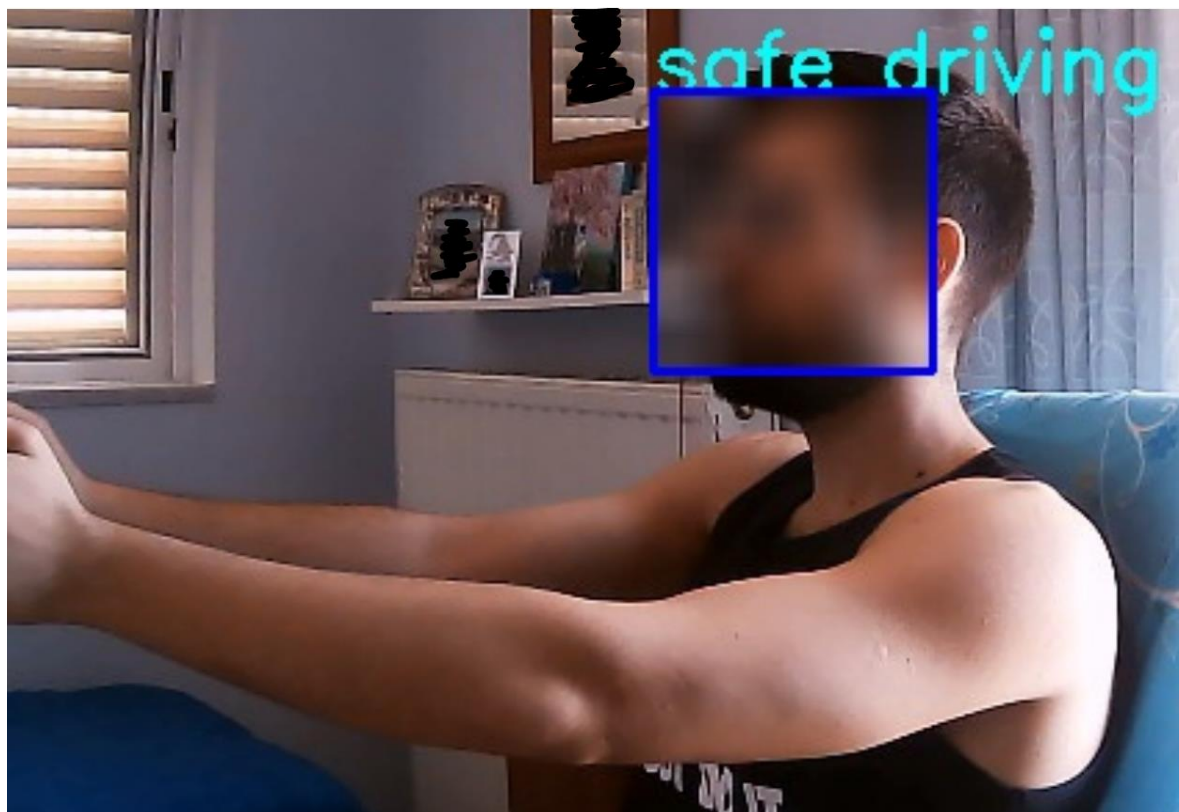


Εικόνα 59: Το confusion matrix του ensemble στο επεξεργασμένο σύνολο δεδομένων.

Όπως ήταν αναμενόμενο, το ensemble των δυο μοντέλων πετυχαίνει καλύτερα αποτελέσματα σε κάθε είδος βαθμολογίας επιδόσεων στο επεξεργασμένο σύνολο δεδομένων.

2.5.7 Πρόβλεψεις σε πραγματικό χρόνο

Τελευταίο και σημαντικό βήμα είναι η μελέτη της συμπεριφοράς του ensemble των μοντέλων σε πραγματικό χρόνο. Η διαδικασία που ακολουθήθηκε μοιάζει με αυτήν για την επεξεργασία του συνόλου των δεδομένων. Ήτοι σε κάθε καρέ της κάμερας εντοπίζεται το πρόσωπο του οδηγού και θολώνεται για λόγους ιδιωτικότητας αλλά και επιδόσεων -για τους λόγους που έχουν αναλυθεί παραπάνω- και ύστερα υπόκειται σε επεξεργασία ώστε να έρθουν στην μορφή που περιμένουν τα μοντέλα. Έπειτα το επεξεργασμένο πλέον καρέ αποστέλλεται στο ensemble των μοντέλων και το αποτέλεσμα της πρόβλεψης εμφανίζεται στην ζωντανή ροή, όπως φαίνεται στην εικόνα παρακάτω.



Εικόνα 60: Πρόβλεψη στάσης οδηγού σε πραγματικό χρόνο

Δυστυχώς δεν υπήρχε η δυνατότητα δοκιμής σε πραγματικό περιβάλλον, δηλαδή μέσα σε αυτοκίνητο. Κατά συνέπεια τα μοντέλα που έχουν εκπαιδευτεί σε φωτογραφίες οδηγών που έχουν ληφθεί μέσα σε αυτοκίνητο, αδυνατούσαν τις περισσότερες φορές να κάνουν σωστή πρόβλεψη. Εδώ είναι πολύ πιθανό να πετύχαινε καλύτερες επιδόσεις το InceptionV3 μοντέλο που εκπαιδεύτηκε σε σύνολο δεδομένων που το παρασκήνιο ήταν κενό (μαύρο). Τέλος τα μοντέλα που εκπαιδεύτηκαν (δηλαδή VGG16, ResNet50 & InceptionV3) έχουν έναν τεράστιο αριθμό παραμέτρων, 138.3εκ, 25.6εκ και 23.8εκ αντίστοιχα, κάτι που τα κάνει πολύ απαιτητικά σε πόρους. Ένα σύστημα (πχ αυτοκινήτου) που θα τρέχει σε πραγματικό χρόνο, δεν θα έχει τους απαραίτητους πόρους για να τους διαθέσει.

Κεφάλαιο 3: Συμπεράσματα και Μελλοντικές Βελτιώσεις

3.1 Συμπέρασμα

Ο αριθμός των ατυχημάτων που προκαλούνται από οδηγούς με περισπασμένη προσοχή συνεχώς αυξάνεται παγκοσμίως. Η λύση που προτείνεται σε αυτήν τη διπλωματική εργασία είναι η χρήση της τεχνητής νοημοσύνης που θα αναγνωρίζει την στάση του οδηγού και θα την κατηγοριοποιεί.

Πιο αναλυτικά δοκιμάστηκαν τρία διαφορετικά συνελκτικά νευρωνικά δίκτυα (convolutional neural networks - CNNs) που είχαν ως βάση τους τα προ-εκπαιδευμένα μοντέλα VGG16, ResNet50 και InceptionV3. Το κάθε ένα από αυτά εκπαιδεύτηκε στο σύνολο δεδομένων ονόματι State Farm Distracted Driver Detection, ωστόσο μόνο στο πρώτο χρησιμοποιήθηκε αυτούσιο, το μοντέλο με βάση το ResNet50 τροφοδοτήθηκε με εικόνες όπου τα πρόσωπα των οδηγών είχαν θολωθεί, ενώ το μοντέλο με βάση το InceptionV3 τροφοδοτήθηκε με τμηματοποιημένες εικόνες όπου φαινόταν μόνο οι φιγούρες των οδηγών. Από τα τρία μοντέλα τα καλύτερα αποτελέσματα τα είχε το δεύτερο (δηλαδή το ResNet50) με $\log \text{loss} = 0.1108$ ενώ το VGG16 και InceptionV3 είχαν $\log \text{loss}$ 0.3254 και 15.2563 αντίστοιχα. Ο λόγος που το μοντέλο InceptionV3 είχε τόσο μεγάλο $\log \text{loss}$ οφείλεται στο γεγονός ότι το μοντέλο που εκπαιδεύτηκε στην τμηματοποίηση δεν είχε πρόσβαση σε δεδομένα σχετικά με το αντικείμενο, ως εκ τούτου οι εικόνες που παρήγαγε ήταν χαμηλής ποιότητας. Τέλος τον ensemble των δύο πρώτων μοντέλων έφερε τα αποτελέσματα $\log \text{loss} = 0.1980$ πάνω στο σύνολο δεδομένων με ακρίβεια 0.976. Το ensemble είχε πολύ καλύτερες επιδόσεις στο σύνολο δεδομένων από τα δύο μοντέλα ξεχωριστά, όπως φαίνεται και από το πίνακάκι με τις μετρήσεις αξιολόγησης.

3.2 Μελλοντικές Αλλαγές και Βελτιώσεις

Στο μέλλον έχοντας περισσότερους υπολογιστικούς πόρους, χρόνο και γνώση, δίνετε να δοκιμαστούν τα ακόλουθα με σκοπό την περαιτέρω βελτίωση των αποτελεσμάτων.

1. Να χρησιμοποιηθεί ο ταξινομητής KNN (K-Nearest Neighbors) για να βρεθούν όλες οι γείτονες μίας εικόνας και στη συνέχεια η τελική πρόβλεψη να υπολογίζεται βάση τον μέσο όρο των πιθανοτήτων αυτών των εικόνων. Αυτή η προσέγγιση θα λειτουργούσε καλά λόγω του υψηλού συσχετισμού μεταξύ των εικόνων.
2. Η δημιουργία ενός συνόλου δεδομένων με εικόνες οδηγών και τις μάσκες τους, όπου η μάσκα οι αντίστοιχες εικόνες με σκιαγραφημένες τις φιγούρες των οδηγών ενώ ο περιβάλλοντας χώρος θα είναι κενός. Κατά αυτόν τον τρόπο θα μπορεί να εκπαιδευτεί αποτελεσματικά ένα μοντέλο στην σωστή τμηματοποίηση εικόνων με οδηγούς.
3. Δοκιμή και άλλων ειδών τμηματοποίησης όπως δέρματος, χεριών και προσώπου.
4. Η χρήση συνάρτησης που θα εκτελούσε αυτόματα την θόλωση στις εικόνες που θα τροφοδοτούνταν στο μοντέλο που λόγω λίγων υπολογιστικών πόρων δεν

χρησιμοποιήθηκε. Επίσης η εκπαίδευση ενός μοντέλου στην αναγνώριση και θόλωση προσώπων, θα μπορούσε να αποτελέσει καλή εναλλακτική.

5. Εκπαίδευση μοντέλων με λιγότερους παραμέτρους (όπως το MobileNetV2 με μόνο 3,5εκ) πάνω στα δύο επεξεργασμένα σύνολα δεδομένων (αυτό με τη θόλωση προσώπων και το τμηματοποιημένο) και την χρήση αυτών για την δημιουργία ενός ensemble που θα μπορεί να αποδίδει ακόμα και σ' ένα σύστημα βασισμένο σε CPU.

Βιβλιογραφία – Αναφορές – Διαδικτυακές Πηγές

- [1] «World Health Organization,» [Ηλεκτρονικό]. Available: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>.
- [2] «National Highway Traffic Safety Administration,» [Ηλεκτρονικό]. Available: <https://www.nhtsa.gov/risky-driving/distracted-driving>.
- [3] «Hardison & Cochran, Attorneys at Law,» [Ηλεκτρονικό]. Available: <https://www.lawyernc.com/car-accident-lawyer/common-causes-car-accidents/distracted-driving/>.
- [4] «Distracted Driving 2013 (NHTSA),» April 2015. [Ηλεκτρονικό]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812132>.
- [5] «Distracted Driving 2014,» April 2016. [Ηλεκτρονικό]. Available: <https://crashstats.nhtsa.dot.gov/api/public/viewpublication/812260>.
- [6] «Distracted Driving 2015,» March 2017. [Ηλεκτρονικό]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812381>.
- [7] «Distracted Driving 2016,» April 2018. [Ηλεκτρονικό]. Available: <https://crashstats.nhtsa.dot.gov/api/public/viewpublication/812517>.
- [8] «Distracted Driving in Fatal Crashes, 2017 (NHTSA),» April 2019. [Ηλεκτρονικό]. Available: <https://crashstats.nhtsa.dot.gov/api/public/viewpublication/812700>.
- [9] «Distracted Driving 2018 (NHTSA),» April 2020. [Ηλεκτρονικό]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812926>.
- [10] «Distracted Driving 2019 (NHTSA),» April 2021. [Ηλεκτρονικό]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813111>.
- [11] «Wikipedia/Advanced driver-assistance systems,» [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Advanced_driver-assistance_systems.
- [12] State Farm Insurance, «Kaggle,» 2016. [Ηλεκτρονικό]. Available: <https://www.kaggle.com/c/state-farm-distracted-driver-detection/overview>.
- [13] F. Chollet, Deep Learning with Python, Second Edition, Manning, 2021.
- [14] «TensorFlow,» Google, [Ηλεκτρονικό]. Available: <https://www.tensorflow.org/>.
- [15] «Wikipedia/TensorFlow,» Wikimedia Foundation, [Ηλεκτρονικό]. Available: <https://en.wikipedia.org/wiki/TensorFlow>.

- [16] F. Chollet, «Keras,» Google, [Ηλεκτρονικό]. Available: <https://keras.io/>.
- [17] «Wikipedia/Keras,» Wikipedia Foundation , [Ηλεκτρονικό]. Available: <https://en.wikipedia.org/wiki/Keras>.
- [18] V. Silaparasetty, «Medium,» 27 July 2019. [Ηλεκτρονικό]. Available: <https://medium.datadriveninvestor.com/how-to-handle-overfitting-and-underfitting-470a1f7389fe>.
- [19] «TensorFlow/tutorials,» Google, [Ηλεκτρονικό]. Available: <https://www.tensorflow.org/tutorials/images/segmentation>.
- [20] «Wikipedia/U-Net,» [Ηλεκτρονικό]. Available: <https://en.wikipedia.org/wiki/U-Net>.
- [21] «Wikipedia/Ensemble_learning,» Wikimedia Foundation, [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Ensemble_learning.