ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Βαθιά Μηχανική Μάθηση

Image Emotion Recognition using CNN

Κούγκουλα Μαγδαληνή

Ταπτά Ελένη

Αθήνα, Ιούλιος 2021

Περίληψη

Καθώς συνεχώς κινούμαστε προς έναν ψηφιακό κόσμο, η Αλληλεπίδραση Ανθρώπου-Υπολογιστή γίνεται πολύ σημαντική, γεγονός που έχει οδηγήσει σε σημαντικές έρευνες σε αυτόν τον τομέα την τελευταία δεκαετία. Οι εκφράσεις προσώπου αποτελούν βασικό χαρακτηριστικό της μη λεκτικής επικοινωνίας και διαδραματίζουν σημαντικό ρόλο στην Αλληλεπίδραση Ανθρώπου Υπολογιστή. Αυτό το έγγραφο παρουσιάζει μια προσέγγιση της αναγνώρισης έκφρασης προσώπου (FER) χρησιμοποιώντας τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Network - CNN). Το μοντέλο που δημιουργήθηκε με τη χρήση του CNN μπορεί να χρησιμοποιηθεί για την ανίχνευση εκφράσεων προσώπου και σε πραγματικό χρόνο. Το σύστημα μπορεί να χρησιμοποιηθεί για την ανάλυση συναισθημάτων ενώ οι χρήστες παρακολουθούν τρέιλερ ταινιών ή διαλέξεις βίντεο. Υπάρχει ανάγκη για μια εφαρμογή που θα είναι σε θέση να ανιχνεύει και να ταξινομεί τις ανθρώπινες εκφράσεις σε πραγματικό χρόνο. Αυτή η ταξινόμηση των συναισθημάτων μπορεί στη συνέχεια να χρησιμοποιηθεί για την κατανόηση του ανθρώπινου νου στον τομέα της ψυχολογίας ή για να βοηθήσει τις μηχανές να κατανοήσουν τις απαιτήσεις των χρηστών.

Η εργασία σκοπεύει να παρουσιάσει μία μέθοδο για την ανάπτυξη ενός αλγορίθμου στο σύνολο δεδομένων FER2013 χρησιμοποιώντας το CNN. Ο αλγόριθμος θα ταξινομήσει την έκφραση ενός ανθρώπινου προσώπου σε μία από τις επτά εκφράσεις – θυμός, ευτυχία, θλίψη, έκπληξη, φόβος, ουδέτερος, αηδία. Το μοντέλο που αναπτύσσεται έτσι μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση των ανθρώπινων προσώπων από φωτογραφίες. Αυτός ο αλγόριθμος μπορεί να χρησιμοποιηθεί για την ανάλυση των εκφράσεων των χρηστών, και να κατανοήσει καλύτερα τις ανθρώπινες απαιτήσεις.

Δεδομένα

Το σύνολο δεδομένων που χρησιμοποιήθηκε για την εφαρμογή του συστήματος ήταν το σύνολο δεδομένων FER2013 από την πρόκληση Kaggle στο FER. Το σύνολο δεδομένων αποτελείται από 35.887 επισημασμένες εικόνες, οι οποίες χωρίζονται σε 3589 train και 28709 test εικόνες. Το σύνολο δεδομένων αποτελείται από άλλες 3589 private test εικόνες, στις οποίες διεξήχθη η τελική δοκιμή κατά τη διάρκεια της πρόκλησης FER. Οι εικόνες στο σύνολο δεδομένων FER2013 έχουν μέγεθος 48x48 και είναι ασπρόμαυρες εικόνες. Το σύνολο δεδομένων FER2013 περιέχει εικόνες που διαφέρουν ως προς την οπτική γωνία, τον φωτισμό και την κλίμακα. Η εικόνα 1 εμφανίζει ορισμένα δείγματα εικόνων από το σύνολο δεδομένων FER2013 και ο πίνακας 1 απεικονίζει την περιγραφή του συνόλου δεδομένων.

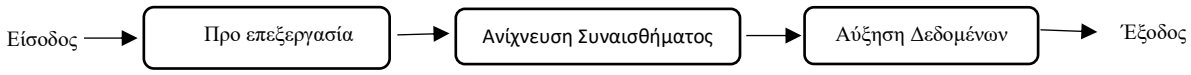
Εικόνα 1.

Αριθμός	Αριθμός Εικόνων	Συναίσθημα
0	4593	θυμός
1	547	αηδία
2	5121	φόβος
3	8989	ευτυχία
4	6077	λύπη
5	4002	έκπληξη
6	6198	ουδέτερος

Πίνακας 1. Περιγραφή του συνόλου δεδομένων FER2013

Διαδικασία Ταξινόμησης

Η διαδικασία έχει τρία στάδια. Το στάδιο της προ-επεξεργασίας έγκειται στην προετοιμασία του συνόλου δεδομένων για να το επεξεργαστεί ο αλγόριθμος και να παράγει αποτελέσματα. Το βήμα ταξινόμησης/ανίχνευσης συναισθημάτων γίνεται κατά την εφαρμογή του δικτύου CNN για την ταξινόμηση της εικόνας εισόδου σε μία από τις επτά κατηγορίες. Τέλος έχουμε την αύξηση δεδομένων. Η αύξηση δεδομένων εικόνας είναι μια τεχνική που μπορεί να χρησιμοποιηθεί για την τεχνητή επέκταση του μεγέθους ενός συνόλου δεδομένων κατάρτισης δημιουργώντας τροποποιημένες εκδόσεις εικόνων στο σύνολο δεδομένων. Η εκπαίδευση μοντέλων νευρωνικών δικτύων βαθιάς μάθησης σε περισσότερα δεδομένα μπορεί να οδηγήσει σε πιο επιδέξια μοντέλα και οι τεχνικές αύξησης μπορούν να δημιουργήσουν παραλλαγές των εικόνων, με τέτοιο τρόπο ώστε να μπορούν να βελτιώσουν την ικανότητα των μοντέλων προσαρμογής να γενικεύσουν ό,τι έχουν μάθει σε νέες εικόνες. Τα στάδια αυτά περιγράφονται με τη χρήση σε διάγραμμα ροής στην εικόνα 2.



Εικόνα 2.

Προ-Επεξεργασία

Η εικόνα εισόδου στο σύστημα μπορεί να περιέχει θόρυβο και να έχει διακυμάνσεις στο φωτισμό, το μέγεθος και το χρώμα. Για να λάβουμε ακριβή και ταχύτερα αποτελέσματα στον αλγόριθμο, έγιναν ορισμένες λειτουργίες προ-επεξεργασίας στην εικόνα. Οι στρατηγικές που χρησιμοποιούνται είναι η ομαλοποίηση, η αλλαγή μεγέθους της εικόνας και η τυποποίηση. Μέσα από τον κώδικα δίνονται ενδεικτικά παραδείγματα 15 τυχαίων εικόνων για τις οποίες έχει εφαρμοστεί η συγκεκριμένη διαδικασία.

1) Ομαλοποίηση (Normalization) - Η ομαλοποίηση μιας εικόνας γίνεται για να αφαιρεθούν οι παραλλαγές φωτισμού και να ληφθεί βελτιωμένη εικόνα προσώπου.

2) Αλλαγή μεγέθους - Η εικόνα αλλάζει μέγεθος για να αφαιρεθούν τα περιττά τμήματα της εικόνας. Αυτό μειώνει την απαιτούμενη μνήμη και αυξάνει την ταχύτητα υπολογισμού.

3) Η τυποποίηση (Standardization) - Είναι μια τεχνική που κλιμακώνει τα δεδομένα, λαμβάνοντας την υπόθεση ότι η κατανομή των δεδομένων είναι Gaussian και μετατοπίζει την κατανομή των δεδομένων σε «zero mean» και «unit (1) standard deviation». Τα δεδομένα με αυτόν τον τύπο διανομής είναι γνωστά ως τυπικά Gaussian.

Οι τυποποιημένες εικόνες προκύπτουν αφαιρώντας τις μέσες τιμές pixel από τις μεμονωμένες τιμές pixel και στη συνέχεια διαιρώντας τις με την τυπική απόκλιση των τιμών pixel. Τα ακόλουθα βήματα πρέπει να ληφθούν για την τυποποίηση εικονοστοιχείων εικόνας:

1. Υπολογισμός της μέσης και τυπικής απόκλισης των τιμών pixel.
2. Χρήση στατιστικών για την τυποποίηση κάθε εικόνας. Στο Keras, αναφέρεται ως τυποποιημένη τυποποίηση.
3. Δημιουργία μιας παρτίδας εικόνων που έχουν μηδενική μέση τυπική απόκλιση μονάδας ώστε το δείγμα να πλησιάζει το τυπικό Gaussian.
4. Εκτέλεση της δοκιμής σε ολόκληρο το σύνολο δεδομένων, προκειμένου να επιβεβαιωθεί ότι ο μέσος όρος είναι κοντά στο μηδέν και η τυπική απόκλιση είναι κοντά στο 1.
5. Εφαρμογή κλιμάκωσης pixel κατά την προσαρμογή και αξιολόγηση του νευρωνικού δικτύου.

<u>Ανίχνευση συναισθημάτων</u>

Σε αυτό το βήμα, το σύστημα ταξινομεί την εικόνα σε μία από τις επτά συναισθηματικές εκφράσεις - Ευτυχία, Θλίψη, Θυμός, Έκπληξη, Αηδία, Φόβος και Ουδέτερο, όπως επισημαίνεται στο σύνολο δεδομένων FER2013. Η εκπαίδευση πραγματοποιήθηκε με τη χρήση του CNN, τα οποία αποτελούν μια κατηγορία νευρωνικών δικτύων που έχουν αποδειχθεί πολύ αποτελεσματικά στην επεξεργασία εικόνας. Το σύνολο δεδομένων χωρίστηκε αρχικά σε σύνολα δεδομένων train και test και στη συνέχεια εκπαιδεύτηκε στο σύνολο train.

Η προσέγγιση που εφαρμόστηκε ήταν να πειραματιστούμε με διαφορετικές αρχιτεκτονικές στο CNN, για να επιτύχουμε όσο γίνεται καλύτερη ακρίβεια με το σύνολο validation, με ελάχιστο overfitting. Το στάδιο ταξινόμησης συναισθημάτων αποτελείται από τις ακόλουθες φάσεις:

1) **Διαχωρισμός δεδομένων**

Το σύνολο δεδομένων χωρίστηκε σε 3 κατηγορίες σε training, test και validation σύνολο δεδομένων.

2) **Εκπαίδευση και παραγωγή μοντέλου**

Η αρχιτεκτονική του νευρικού δικτύου αποτελείται από τα ακόλουθα επίπεδα:

1. ***Στρώμα συνένωσης (Convolution Layer)***
   Στο επίπεδο συνένωσης, ένα τυχαία στιγμιαίο instantiated learnable filter (φίλτρο) γλιστράει (slide) ή περιστρέφεται πάνω από την είσοδο. Η λειτουργία εκτελεί το εσωτερικό γινόμενο μεταξύ του φίλτρου και κάθε τοπικής περιοχής της εισόδου. Η έξοδος είναι ένας τρισδιάστατος τόμος πολλαπλών φίλτρων, που ονομάζεται activation map.

2. ***Μέγιστη ομαδοποίηση (Max Pooling)***
   Η μέγιστη ομαδοποίηση χρησιμοποιείται για τη μείωση του χωρικού μεγέθους του activation map για τη μείωση του μεγέθους της εισόδου και του κόστους υπολογισμού.

3. ***Πλήρως συνδεδεμένο επίπεδο (Fully Connected Layer)***
   Στο πλήρως συνδεδεμένο στρώμα, κάθε νευρώνας από το προηγούμενο στρώμα συνδέεται με τους νευρώνες εξόδου. Το μέγεθος του τελικού output layer ισούται με τον αριθμό των κλάσεων στις οποίες πρόκειται να ταξινομηθεί η εικόνα εισόδου.

4. ***Συνάρτηση ενεργοποίησης (Activation function)***
   Οι συναρτήσεις ενεργοποίησης χρησιμοποιούνται για τη μείωση του overfitting. Στην αρχιτεκτονική CNN, έχει χρησιμοποιηθεί η λειτουργία ενεργοποίησης ReLu. Το πλεονέκτημα της συνάρτησης ενεργοποίησης ReLu είναι ότι η κλίση της είναι πάντα ίση με 1, πράγμα που σημαίνει ότι το μεγαλύτερο μέρος του σφάλματος μεταβιβάζεται πίσω κατά την πίσω διάδοση.

$$f(x) = \max(0, x)$$

5. ***Μαλακή κορύφωση (Softmax)***
   Η συνάρτηση softmax παίρνει ένα διάνυσμα των πραγματικών αριθμών N και ομαλοποιεί αυτό το διάνυσμα σε μια σειρά τιμών μεταξύ (0, 1).

6. ***Κανονικοποίηση παρτίδας (Batch Normalization)***
   Ο ομαλοποιητής παρτίδας επιταχύνει τη διαδικασία εκπαίδευσης και εφαρμόζει έναν μετασχηματισμό που διατηρεί τη μέση ενεργοποίηση κοντά στο 0 και την τυπική απόκλιση ενεργοποίησης κοντά στο 1.

7. ***Αξιολόγηση του μοντέλου***
   Το μοντέλο που δημιουργήθηκε κατά τη διάρκεια της φάσης εκπαίδευσης αξιολογήθηκε στη συνέχεια στο validation set, το οποίο αποτελούνταν από 3589 εικόνες.

Αύξηση Δεδομένων (Data Augmentation)

Η αύξηση δεδομένων είναι μια τεχνική που μπορεί να χρησιμοποιηθεί για την τεχνητή επέκταση του μεγέθους ενός συνόλου δεδομένων training δημιουργώντας τροποποιημένες εκδόσεις εικόνων στο σύνολο δεδομένων. Η εκπαίδευση μοντέλων νευρωνικών δικτύων βαθιάς μάθησης σε περισσότερα δεδομένα μπορεί να οδηγήσει σε πιο επιδέξια μοντέλα και οι τεχνικές αύξησης μπορούν να δημιουργήσουν παραλλαγές των εικόνων που μπορούν να βελτιώσουν την ικανότητα των μοντέλων προσαρμογής να γενικεύσουν αυτό που έχουν μάθει σε νέες εικόνες. Η βιβλιοθήκη νευρωνικών δικτύων βαθιάς μάθησης Keras παρέχει τη δυνατότητα προσαρμογής μοντέλων χρησιμοποιώντας αύξηση δεδομένων εικόνας μέσω της συνάρτησης ImageDataGenerator .

Αύξηση δεδομένων εικόνας

Η απόδοση των νευρωνικών δικτύων βαθιάς μάθησης βελτιώνεται συχνά με την ποσότητα των διαθέσιμων δεδομένων. Η αύξηση δεδομένων είναι μια τεχνική για την τεχνητή δημιουργία νέων δεδομένων εκπαίδευσης από υπάρχοντα εκπαιδευτικά δεδομένα. Αυτό επιτυγχάνεται με την εφαρμογή τεχνικών συγκεκριμένων τομέων σε παραδείγματα από τα δεδομένα εκπαίδευσης που δημιουργούν νέα και διαφορετικά παραδείγματα εκπαίδευσης. Η αύξηση δεδομένων εικόνας είναι ίσως ο πιο γνωστός τύπος αύξησης δεδομένων και περιλαμβάνει τη δημιουργία μετασχηματισμένων εκδόσεων εικόνων στο σύνολο δεδομένων εκπαίδευσης που ανήκουν στην ίδια τάξη με την αρχική εικόνα. Οι μετασχηματισμοί περιλαμβάνουν μια σειρά λειτουργιών από το πεδίο της επεξεργασίας (manipulation) της εικόνας, όπως μετατοπίσεις, αναστροφές, ζουμ και πολλά άλλα. Ο σκοπός είναι να επεκταθεί το σύνολο δεδομένων

εκπαίδευσης με νέα, εύλογα παραδείγματα. Αυτό σημαίνει, παραλλαγές των εικόνων που καθορίζονται από το μοντέλο. Για παράδειγμα, ένα οριζόντιο γύρισμα μιας εικόνας μιας γάτας μπορεί να έχει νόημα, επειδή η φωτογραφία θα μπορούσε να έχει τραβηχτεί από αριστερά ή δεξιά. Ένα κατακόρυφο γύρισμα της φωτογραφίας μιας γάτας δεν έχει νόημα και πιθανότατα δεν θα ήταν κατάλληλο δεδομένου ότι το μοντέλο είναι πολύ απίθανο να δει μια φωτογραφία μιας αναποδογυρισμένης γάτας. Ως εκ τούτου, είναι σαφές ότι η επιλογή των συγκεκριμένων τεχνικών αύξησης δεδομένων που χρησιμοποιούνται για ένα σύνολο δεδομένων εκπαίδευσης πρέπει να επιλεγεί προσεκτικά και εντός του πλαισίου του συνόλου δεδομένων κατάρτισης και της γνώσης του προβληματικού τομέα. Επιπλέον, μπορεί να είναι χρήσιμος ο πειραματισμός με μεθόδους αύξησης δεδομένων μεμονωμένα και σε συνδυασμό για να δείτε αν οδηγούν σε μετρήσιμη βελτίωση της απόδοσης του μοντέλου, ίσως με ένα μικρό σύνολο δεδομένων πρωτότυπου, μοντέλου και εκπαίδευσης. Οι σύγχρονοι αλγόριθμοι βαθιάς μάθησης, όπως το συνελικτικό νευρικό δίκτυο, ή το CNN, μπορούν να μάθουν χαρακτηριστικά που είναι αναλλοίωτα στη θέση τους στην εικόνα. Παρ 'όλα αυτά, η αύξηση μπορεί να βοηθήσει περαιτέρω σε αυτήν την μεταβαλλόμενη προσέγγιση της μάθησης και μπορεί να βοηθήσει το μοντέλο σε χαρακτηριστικά εκμάθησης που είναι επίσης αμετάβλητα σε μετασχηματισμούς όπως παραγγελία από αριστερά προς τα δεξιά προς τα πάνω προς τα κάτω, επίπεδα φωτός στις φωτογραφίες και πολλά άλλα. Η αύξηση δεδομένων εικόνας εφαρμόζεται συνήθως μόνο στο σύνολο δεδομένων εκπαίδευσης και όχι στο σύνολο δεδομένων επικύρωσης ή δοκιμής. Αυτό διαφέρει από την προετοιμασία δεδομένων, όπως αλλαγή μεγέθους εικόνας και κλιμάκωση εικονοστοιχείων. Πρέπει να εκτελούνται με συνέπεια σε όλα τα σύνολα δεδομένων που αλληλεπιδρούν με το μοντέλο.

Πειράματα και αποτελέσματα

Τα αποτελέσματα επιτεύχθηκαν με τον πειραματισμό με το δίκτυο CNN. Παρατηρήθηκε ότι η απώλεια (loss) κατά τη διάρκεια του train και του συνόλου test μειώθηκε με κάθε epoch. Το μέγεθος της παρτίδας (batch size) ήταν 256, το οποίο παρέμεινε σταθερό σε όλα τα πειράματα.

Έγιναν οι ακόλουθες αλλαγές στην αρχιτεκτονική του νευρικού δικτύου για την επίτευξη καλών αποτελεσμάτων:

1) Αριθμός επαναλήψεων (epoch):

   Παρατηρήθηκε ότι η ακρίβεια του μοντέλου αυξήθηκε με αυξανόμενο αριθμό επαναλήψεων. Ωστόσο, ένας μεγάλος αριθμός επαναλήψεων είχε ως αποτέλεσμα την υπερπροσφορά. Συνήχθη το συμπέρασμα ότι οκτώ εποχές κατέληξαν σε ελάχιστη υπερπροσφορά και υψηλή ακρίβεια.

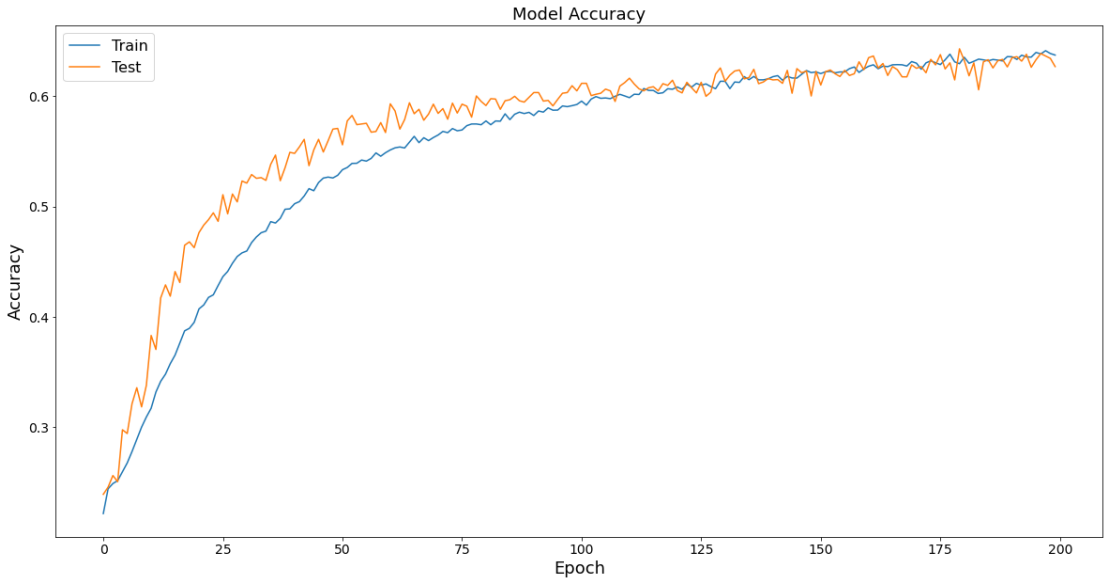
2) Αριθμός στρώσεων:

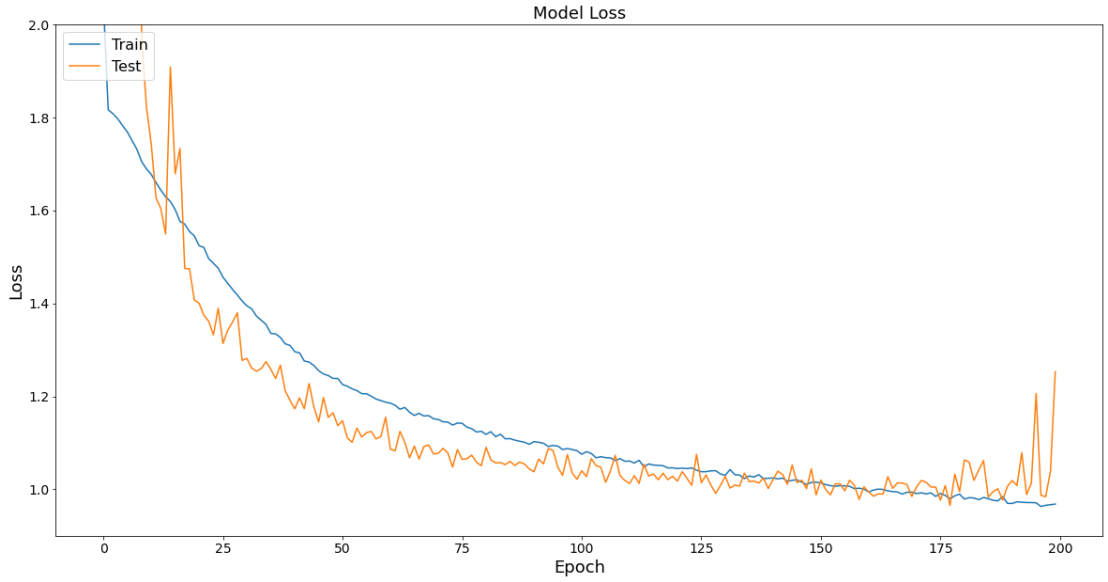
   Η αρχιτεκτονική του νευρικού δικτύου αποτελείται από τρία κρυμμένα στρώματα και ένα ενιαίο πλήρως συνδεδεμένο στρώμα. Κατασκευάστηκαν συνολικά έξι επίπεδα συνένωσης, χρησιμοποιώντας τη «RelU» ως συνάρτηση ενεργοποίησης (activation function).

3) Φίλτρα:
   Η ακρίβεια του νευρωνικού δικτύου στο σύνολο δεδομένων ποικίλλει ανάλογα με τον αριθμό των φίλτρων που εφαρμόζονται στην εικόνα. Ο αριθμός των φίλτρων για τα δύο πρώτα επίπεδα του δικτύου ήταν 64 και 128 και 256 για τα δύο τελευταία επίπεδα του δικτύου.

Απώλεια και ακρίβεια με την πάροδο του χρόνου

Μπορεί να παρατηρηθεί ότι η απώλεια μειώνεται και η ακρίβεια αυξάνεται με κάθε epoch. Η καμπύλη εκπαίδευσης (Train), έναντι της δοκιμής (Test), αναφορικά με το επίπεδο της ακρίβειας (Accuracy) παραμένει υψηλή μετά από τα πρώτα 100 epochs, σε αντίθεση με τα πρώτα 100, όπου αποκλίνει από τις ιδανικές τιμές. Η ακρίβεια της εκπαίδευσης και των δοκιμών μαζί με την απώλεια εκπαίδευσης και επικύρωσης που αποκτήθηκαν για το σύνολο δεδομένων FER2013 με τη χρήση του CNN παρατίθενται στον παρακάτω πίνακα:





Παρακάτω βλέπουμε εικόνες που έχουν προβλεφθεί σωστά και εικόνες που έχουν προβλεφθεί λάθος.

Το confusion matrix που δημιουργείται μέσω των δεδομένων δοκιμής παρουσιάζεται στο παρακάτω σχήμα. Τα σκοτεινά μπλοκ κατά μήκος της διαγώνιας δείχνουν ότι τα δεδομένα δοκιμής έχουν ταξινομηθεί καλά. Μπορεί να παρατηρηθεί ότι ο αριθμός των σωστών

ταξινομήσεων είναι χαμηλός για την κλάση «Disgust» (Αηδία), ακολουθούμενος από εκείνες για την κλάση «Fear» (Φόβος). Οι αριθμοί δεξιά και αριστερά της διαγωνίου αντιπροσωπεύουν τον αριθμό των εσφαλμένα ταξινομημένων εικόνων. Δεδομένου ότι αυτοί οι αριθμοί είναι χαμηλότεροι σε σύγκριση με τους αριθμούς στη διαγώνιο, μπορεί να προκύψει το συμπέρασμα ότι ο αλγόριθμος λειτούργησε σωστά και πέτυχε ικανοποιητικά αποτελέσματα.