Διάβασα το paper, και κοίταξα λίγο το github repostitory στο οποίο παραθέτει τον κώδικα.

Στο paper κάνει λόγω και για machine learning που εμας δεν μας αφορά, επίσης φαντάζομαι ότι το μοντέλο του νευρωνικού δικτύου δεν θα είναι απλά ένα βαθύ νευρωνικό δίκτυο που θα πρέπει να μάθει μόνο του να κάνει extract τα features αλλά θα του προσφέρουμε εμείς κάποια engineered features για αρχή.

Έχω κάποιες απορίες:

Α) Μέσα στο paper τα δίκτυα που εκπαιδεύει αυτός έχουνε σαν είσοδο features από text το (transcript) και ήχο. Δηλαδή αγνοεί την πληροφορία για την εικόνα, τουλάχιστον αυτό κατάλαβα.  
Εγώ στο δικό μου δίκτυο όμως θα έχω σαν είσοδο features από εικόνα και ήχο σωστά;; Όχι κείμενο και ήχο.  
Άρα από το paper αυτό κρατάω μόνο τους τρόπους επεξεργασίας του ήχου για την δικιά μου επίλυση.

A close up of a text

AI-generated content may be incorrect.

Το dataset που θα δουλέψω εγώ το MELD αφορά διαλόγους πολλών ανθρώπων και όχι συζήτηση μεταξύ δύο ατόμων όπως είναι το IEMOCAP (δυαδικής φύσης) με το οποίο δουλεύει το paper αυτό. Έχει επίσης για κάθε διάλογο ήχο, εικόνα και text ενώ το IEMOCAP έχει μόνο text+audio άρα δεν μπορώ να χρησιμοποιήσω τον πίνακα που κάνει rank τα διαφορετικά μοντέλα, γιατί το δικό μου δίκτυο θα κατηγοριοποιεί με βάση την εικόνα και τον ήχο και αυτουνού κατηγοριοποιεί με βάση τον ήχο και τα λεγόμενα.

A table with numbers and text

AI-generated content may be incorrect.

A2) Και στο MELD με ήχο και text features πάνε αλλά νομίζω βάζουν και εικόνα;;;

Α3) Τι είναι τα baseline models.

A4) Για να βγάλω στοιχεία από την εικόνα θα πρέπει κάθε φορά να ανιχνεύω σωστά την φάτσα του ομιλητή και μετά να βρίσκω τι νιώθει, γιατί σε κάθε διάλογο βρίσκονται πολλοί άνθρωποι και κάθε φορά μιλάει διαφορετικός.

Β) Άρα πρέπει να ψάξω για δίκτυα που παίρνουν εικόνα και ήχο και κατηγοριοποιούν τα συναισθήματα. Θα ψάξω αυτά τα papers που είδα στο github του dataset μου ότι το κάνουν reference μπας και δουλεύουν με εικόνα και ήχο αυτά. (Δεν τα έχω δει ακόμα ούτε τα έχω ψάξει ελπίζω να τα βρω)

A screenshot of a black screen

AI-generated content may be incorrect.

Γ) Που μπορώ να βρίσκω τα papers που κάνει reference διότι όταν πάω από κάτω στο άρθρο και αντιγράφω τους τίτλους μαζί με τους επιστήμονες δεν τα βρίσκω στο google και έψαχνα το 5 άρθρο που εξηγούσε το database που χρησιμοποιεί και δεν το βρήκα. To βρήκα στο google scholar αλλά κόστιζε χρήματα.

Δ) Γενικά εμένα με νοιάζει σαν μετρική να έχουνε καλό F1 σκορ ,δηλαδή να είναι το ίδιο σημαντικό το precision και το recall ή μπορεί ένα από αυτά να μην μας πειράζει τόσο και να δίνουμε προτεραιότητα στο precision ή στο recall ανάποδα.

Φ1 καλο

Ε) Την αναφορά θα την κάνω σε latex σωστά, είναι καλό να βάλω στην διπλωματική ένα μέρος που να περιγράφει την έρευνα που κάνουμε τώρα ,πού έψαξα κτλ ή δεν υπάρχει λόγος.  
Να αρχίσω από τώρα δηλαδή να γράφω ότι έψαξα εκεί και εδώ αλλά αυτό δεν έκανε επειδή τάδε κτλ

Πρώτο κεφάλαιο το πρόβλημα multimodal emontional network

Δευτερο κεφάλαιο το πρόβλημα συγκεκριμένα multimodal emontional network text και audio, περιγράφω state of the art μοντέλο.

Τρίτο κεφάλαιο η λύση μου.

Τέταρτο κεφάλαιο αποτελέσματα και σύγκριση με άλλα modela άλλων.