

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Σίμου Νίκων
Μεταπτυχιακός Φοιτητής**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Π. Τσακαλίδης**

**Τετάρτη, 09/10/2019, 09:00
Αίθουσα Β106, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**“ Εκτίμηση ποιότητας ηχογραφήσεων από μουσικές συναυλίες με χρήση τεχνικών
βαθιάς μάθησης”**

ΠΕΡΙΛΗΨΗ

Η εποχή στην οποία ζούμε χαρακτηρίζεται αδιαμφισβήτητα από ένα μεγάλο όγκο πολυμεσικής πληροφορίας. Χρησιμοποιώντας συσκευές όπως τα smartphones και τα drones μπορούμε να απαθανατίσουμε κάθε στιγμή της ζωής μας και των εκδηλώσεων στις οποίες παρευρισκόμαστε. Το υλικό που καταγράφεται από δημόσιες εκδηλώσεις συχνά γίνεται διαθέσιμο στο ευρύ κοινό μέσα από μέσα κοινωνικής δικτύωσης καθώς και από ιστοσελίδες που παρέχουν οπτικοακουστικό περιεχόμενο. Η διαθεσιμότητα του υλικού αυτού, έχει δημιουργήσει νέες ερευνητικές τάσεις σχετικά με την αναζήτηση, την οργάνωση και την επεξεργασία του.

Στην παρούσα εργασία χρησιμοποιούμε τεχνικές Βαθιάς Μάθησης με σκοπό την δημιουργία ενός εργαλείου το οποίο θα μπορεί να εκτιμά αυτόματα την ποιότητα ηχογραφήσεων που παράγονται από χρήστες που παρακολουθούν και καταγράφουν

μουσικές συναυλίες. Όπως είναι ευρέως γνωστό οι τεχνικές Βαθιάς Μάθησης απαιτούν έναν πολύ μεγάλο αριθμό δεδομένων εκπαίδευσης, γεγονός που θα απαιτούσε τη συλλογή υποκειμενικών εκτιμήσεων από ένα μεγάλο όγκο ηχητικών δεδομένων. Προκειμένου να ξεπεράσουμε τη δυσκολία αυτή, αντιμετωπίζουμε την εκτίμηση ποιότητας ως ένα πρόβλημα δυαδικής ταξινόμησης όπου η κλάση 0 αντιπροσωπεύεται από ηχογραφήσεις που παράγονται από χρήστες, ενώ η κλάση 1 αντιπροσωπεύεται από το επαγγελματικό ηχητικό υλικό της μουσικής συναυλίας. Επιπλέον, χρησιμοποιούμε μια αυτόματη διαδικασία συγχρονισμού με σκοπό κάθε ηχογράφιση προερχόμενη από χρήστη να εισέρχεται ως δεδομένο εκπαίδευσης μαζί με το αντίστοιχο απόσπασμα από το επαγγελματικό υλικό. Η διαδικασία αυτή συνεισφέρει έτσι ώστε η εκμάθηση του συστήματος να βασιστεί όσο το δυνατόν περισσότερο σε ποιοτικά χαρακτηριστικά και να μην εξαρτάται από το μουσικό περιεχόμενο. Παρουσιάζουμε πειραματικά αποτελέσματα με διαφορετικές αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων και διαφορετικά ακουστικά χαρακτηριστικά τα οποία δείχνουν πως ο διαχωρισμός των δύο κλάσεων μπορεί να γίνει με υψηλή ακρίβεια.

Nikon Simou
M.Sc. Thesis

Computer Science Department
University of Crete
Master's Thesis Supervisor: Professor, P. Tsakalides

Wednesday, 09/10/2019, 09:00
Room B106, Computer Science Dept., University of Crete

“Assessing the Quality of Audio in Musical Concert Recordings using Deep Neural Networks”

ABSTRACT

The era in which we live in can be indisputably characterized by the enormous flow of multimedia information. Using portable multimedia devices such as drones and smartphones we are able to capture every moment of our lives and of the public events that we attend. A large proportion of audiovisual recordings from these events becomes available through the social media and the large number of websites which provide video

and audio content. The availability of such massive amount of User Generated Recordings (UGRs) has triggered new research directions related to the search, organization and management of this content.

In this Thesis, we use Deep Neural Networks (DNN) in order to create a tool to automatically assess the audio quality of musical concert recordings that users upload on multimedia platforms such as YouTube. It is well known that DNNs require a lot of training samples, which means that one would need an enormous amount of time in order to listen and to assign a subjective quality score to each audio sample. We tackle this problem by treating quality assessment as a binary classification problem where class 0 consist of the set of UGRs from a certain event and class 1 consist of the professional quality recordings from the same event. Furthermore, we use an automatic synchronization process in order to match every UGR with its corresponding segment from a professional quality recording, which assists in making the process invariant to audio content. Experiments produced with different DNN architectures and acoustic feature are presented, showing that the UGR class can be discriminated from the professional quality class with a high accuracy.