

Ε.Μ.Π. - Σχολή Ηλεκτρολόγων Μηχ/κών & Μηχ/κών Υπολογιστών

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ¹ 2023-2024

Καθηγητής Πέτρος Μαραγκός, E-mail: maragos@cs.ntua.gr
Εργαστήριο: CVSP: <http://cvsp.cs.ntua.gr>, IRAL: <http://robotics.ntua.gr>

Οι ενδιαφερόμενοι/ες παρακαλούνται να αποστείλουν ηλεκτρονικά σε ένα Zip με το όνομά τους τα εξής:
i) την αναλυτική τους βαθμολογία από ΣΗΜΜΥ-ΕΜΠ με μια εκτίμηση του τρέχοντος μέσου όρου, ii) το βιογραφικό τους, και iii) τον αριθμό των οφειλομένων μαθημάτων αν βρίσκονται στο 4^ο ή ανώτερο έτος σπουδών. Η αποστολή της πληροφορίας αυτής να γίνεται στην Γραμματέα του Εργαστηρίου κα. Βίκυ Πλατίτσα (email: <vickyplatitsa@gmail.com>).

Προαπαιτούμενα μαθήματα (Ροής Σ) με επιτυχή και συνεπή παρακολούθηση και πολύ καλή απόδοση:

- Για θέματα N: ΨΕΣ, και (*) Μηχανική Μάθηση-MM ή Αναγνώριση Προτύπων-ΑΠ .
- Για θέματα V: Οραση Υπολογιστών (ΟΥ), ή πιθανώς και MM ή ΑΠ (*).
- Για θέματα R που συνδυάζουν Ρομποτική με CV/SP/ML: ΟΥ και Ρομποτική I ή MM/ΑΠ (*).
- Για θέματα AVM: ΨΕΣ, ΟΥ και MM/ΑΠ (*).
- Για θέματα L: ΨΕΣ, ΟΥ και MM ή Αναγνώριση Προτύπων (*).
- Για θέματα S: ΨΕΣ και κάποιο μάθημα ΣΑΕ απο Ροη Σ, ή πιθανώς και MM ή Αναγνώριση Προτύπων (*).
- Για θέματα T: ΟΥ και MM/ΑΠ (*).

* = μπορεί να παρακολουθείται ταυτόχρονα με διπλωματική.

- Γενικά, για να υπάρχει ποικιλία ερευνητικών θεμάτων, δίνονται περισσότερα θέματα από τις θέσεις που μπορούν να επιβλεφθούν εντός ενός έτους. **Τα ενεργά θέματα εντός του 2023-2024 είναι αυτά με το σήμα (*) στον τίτλο τους.**
- Μετά την εκδήλωση ενδιαφέροντος οι ενδιαφερόμενοι φοιτητές θα συναντηθούν μαζί με τον υπεύθυνο καθ. Π. Μαραγκό και Επιστημονικούς/ές Συνεργάτες/ιδες καθώς και Υποψήφιους/ες Διδάκτορες της ερευνητικής ομάδας και θα γίνει προσπάθεια να αντιστοιχισθούν θέματα και φοιτητές με όσο το δυνατόν καλύτερο ταίριασμα ενδιαφερόντων και ικανοτήτων.
- Η τελική αποδοχή αίτησης για εκπόνηση διπλωματικής θα εξαρτηθεί από την επίδοση στα σχετικά μαθήματα του εργαστηρίου, τον αριθμό των οφειλομένων μαθημάτων, και την χρονική διαθεσιμότητα του θέματος.

Physiological Data Processing and Learning from Wearable Sensors for Healthcare Applications:

Η τεχνολογική πρόοδος που έχει σημειωθεί τα τελευταία χρόνια στις φορητές συσκευές (κινητά τηλέφωνα, smart-watch, κτλ.) έχει πλέον επιτρέψει την αξιόπιστη καταγραφή πολλών και διαφορετικών βιομετρικών δεικτών, καθώς επίσης και ακουστικών σημάτων, τα οποία πλέον χρησιμοποιούνται και στην ιατρική.

(*) N1. Ανάλυση χρονοσειρών βιομετρικών δεικτών για εντοπισμό και πρόβλεψη υποτροπών σε ασθενών με ψυχωτικές διαταραχές

Στόχος της παρούσας διπλωματικής είναι η ανάπτυξη έξυπνων αλγορίθμων που θα εντοπίζουν και θα προβλέπουν υποτροπές σε ασθενείς με ψυχωτικές διαταραχές μέσω της αξιοποίησης δεδομένων που καταγράφονται από έξυπνες-φορητές συσκευές κατά τη διάρκεια της καθημερινότητας των ασθενών. Η αξιοποίηση της μηχανικής μάθησης σε βιομετρικά δεδομένα μπορεί να δώσει το συμπεριφορικό προφίλ των ασθενών και στη συνέχεια να εντοπίσει αποκλίνουσες συμπεριφορές από το σύνηθες μοτίβο. Έτσι, είναι δυνατό να εντοπιστούν μεταβολές και τάσεις που θα αποκαλύψουν μια ψυχωτική υποτροπή. Η πρόβλεψη μιας τέτοιας υποτροπής μπορεί να αποτελέσει ένα σημαντικό εργαλείο στην έγκαιρη αντιμετώπιση της.

(Συνεργάτες: ΥΔ Νίκη Ευθυμίου, Δρ. Γιώργος Ρετινιάς, Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ, και Καθ. Ν. Σμυρνής, ΕΚΠΑ & ΕΠΙΨΥ)

¹ Μερικά από τα ανωτέρω θέματα έχουν προοπτική για Διδακτορικό με οικονομική υποστήριξη από ερευνητικά προγράμματα.

References:

- [1] N. Efthymiou, G. Retsinas, P. Filntisis, C. Garoufis, A. Zlatintsi, E. Kalisperakis, V. Garyfalli, T. Karantinos, M. Lazaridi, N. Smyrnis, and P. Maragos, From digital phenotype identification to detection of psychotic relapses, in Proc. IEEE International Conference on Health Informatics (ICHI), 2023.
- [2] E. Fekas, A. Zlatintsi, P. Filntisis, C. Garoufis, N. Efthymiou, and P. Maragos, [Relapse prediction from long-term wearable data using self-supervised learning and survival analysis](#), in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023
- [3] G. Retsinas, P.P. Filntisis, N. Efthymiou, E. Theodosios, A. Zlatintsi and P. Maragos, [Person Identification Using Deep Convolutional Neural Networks On Short-term Signals From Wearable Sensors](#), in Proc. 45th IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP-2020), May 2020.
- [4] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, "[E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures](#)," Sensors, vol. 22, no. 19, Oct. 2022.
- [5] C. Lee & M. Van Der Schaar, [Temporal phenotyping using deep predictive clustering of disease progression](#). In Proc. International Conference on Machine Learning (ICML), 2020.
- [6] D. A. Adler et al., [Predicting Early Warning Signs of Psychotic Relapse From Passive Sensing Data: An Approach Using Encoder-Decoder Neural Networks](#), JMIR mHealth and uHealth, Vol. 8, 2020.

(*) N2. Ανάλυση χρονοσειρών βιομετρικών δεικτών και βίντεο δεδομένων σε εφαρμογές ψυχικής υγείας

Ένα σύγχρονο πεδίο έρευνας που αναπτύσσεται και επεκτείνεται διαρκώς είναι αυτό της ηλεκτρονικής υγείας (eHealth). Η αυξανόμενη χρήση wearable και audiovisual συσκευών έχει ως αποτέλεσμα τη συλλογή μεγάλου όγκου πολυμεσικών δεδομένων που καταγράφουν πολλές πτυχές της καθημερινότητάς των ανθρώπων. Ο συνδυασμός αυτών των δεδομένων με την αλματώδη ανάπτυξη της μηχανικής μάθησης έχουν ανοίξει νέες κατευθύνσεις στην ανάλυση σημάτων από πολλαπλούς αισθητήρες για την εκτίμηση της ψυχικής-συναισθηματικής κατάστασης του ατόμου. Στην κατεύθυνση αυτή, η διπλωματική στοχεύει στην ανάπτυξη αλγορίθμων που αξιοποιούν τόσο βιοσήματα όσο και οπτική πληροφορία με σκοπό τον εντοπισμό μη ψυχωτικών υποτροπών (π.χ. καταθλιπτικές υποτροπές).

(Συνεργάτες: ΥΔ Νίκη Ευθυμίου, Δρ. Γιώργος Ρετσίνας, Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] N. Efthymiou, G. Retsinas, P. Filntisis, C. Garoufis, A. Zlatintsi, E. Kalisperakis, V. Garyfalli, T. Karantinos, M. Lazaridi, N. Smyrnis, and P. Maragos, From digital phenotype identification to detection of psychotic relapses, in Proc. IEEE International Conference on Health Informatics (ICHI), 2023.
- [2] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, "[E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures](#)," Sensors, vol. 22, no. 19, Oct. 2022.
- [3] Giannakakis, G., Koujan, M.R., Roussos, A. et al. [Automatic stress analysis from facial videos based on deep facial action units recognition](#). *Pattern Anal Applic* 25, 521–535 (2022).
- [4] Pedrelli, P., Fedor, S., Ghandeharioun, A., Howe, E., Ionescu, D.F., Bhathena, D., Fisher, L.B., Cusin, C., Nyer, M., Yeung, A. and Sangermano, L. [Monitoring changes in depression severity using wearable and mobile sensors](#). *Frontiers in psychiatry*, 11, p.584711.

N3. Ανάλυση ηχητικών σημάτων για υποστήριξη ασθενών σε θέματα ψυχικής υγείας.

Αντικείμενο αυτής της διπλωματικής εργασίας είναι ο εντοπισμός, ή και η πρόβλεψη, υποτροπιαζουσών καταστάσεων σε ασθενείς στο διπολικό ή στο σχιζοειδές φάσμα χρησιμοποιώντας βασιζόμενοι σε δεδομένα ομιλίας των ασθενών κατά τη διάρκεια σύντομων κλήσεων με θεράποντες ιατρούς [1-5]. Ενδεικτικές κατευθύνσεις αποτελούν η χρήση παράλληλων τροπικοτήτων σε συνδυασμό με την ομιλία, όπως για παράδειγμα καταγραφές κειμένου των συνομιλιών [6], και εκμετάλλευση προεκπαιδευμένων γλωσσικών/ακουστικών μοντέλων για το σκοπό αυτό [7,8].

(Συνεργάτες: ΥΔ Χρήστος Γαρούφης, ΥΔ Νίκη Ευθυμίου, Δρ. Νάνσυ Ζλατίντση, ΕΜΠ, και Καθ. Ν. Σμυρνής, ΕΚΠΑ & ΕΠΙΨΥ)

- [1] A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos, A. Menychtas, I. Maglogiannis, P. Tsanakas, T. Sounapoglou, E. Kalisperakis, T. Karantinos, M. Lazaridi, V. Garyfalli, A. Mantas, L. Mantonakis, and N. Smyrnis, “[E-prevention: Advanced support system for monitoring and relapse prevention in patients with psychotic disorders analyzing long-term multimodal data from wearables and video captures](#),” *Sensors*, vol. 22, no. 19, Oct. 2022.
- [2] Garoufis, C., Zlatintsi, A., Filntisis, P. P., Efthymiou, N., Kalisperakis, E., Karantinos, T., ... & Maragos, P. [Towards unsupervised subject-independent speech-based relapse detection in patients with psychosis using variational autoencoders](#). In Proc. EUSIPCO 2022
- [3] J. Gideon, K. Matton, S. Anderau, M.G. McInnis, and E.M. Provost, “[When to Intervene: Detecting Abnormal Mood using Everyday Smart-phone Conversations](#),” arXiv preprint arXiv:1909.11248, 2019.
- [4] K.-Y. Huang, C.-H. Wu, and M.-H. Su, “[Attention-Based Convolutional Neural Network and Long Short-Term Memory for Short-Term Detection of Mood Disorders Based on Elicited Speech Responses](#),” *Pattern Recognition*, vol. 88, pp. 668–678, 2019.
- [5] Z. Pan, C. Gui, J. Zhang, J. Zhu, and D. Cui, “[Detecting Manic State of Bipolar Disorder Based on Support Vector Machine and Gaussian Mixture Model using Spontaneous Speech](#),” *Psychiatry investigation*, vol. 15, no. 7, pp. 695, 2018.
- [6] Yoon, S., Dey, S., Lee, H., & Jung, K. [Attentive modality hopping mechanism for speech emotion recognition](#). In *Proc. ICASSP 2020*
- [7] Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., ... & Wei, F. [Wavlm: Large-scale self-supervised pre-training for full stack speech processing](#). *IEEE Journal of Selected Topics in Signal Processing*, 16(6), 1505-1518.
- [8] Morais, E., Hoory, R., Zhu, W., Gat, I., Damasceno, M., & Aronowitz, H. [Speech emotion recognition using self-supervised features](#). In Proc. ICASSP 2022

(* N4: Multimodal Analysis of Neural and Contextual Signals for Affect Understanding

Current research [1] suggests that complex human constructs like emotions can be reliably comprehended only with regard to the respective personal and social context. This thesis proposal aims to investigate the potential of multimodal signal processing in affective computing by combining neural signals with contextual modalities. The primary objective will be to develop a comprehensive framework that correlates and integrates these data sources [2, 3] to enhance the accuracy and robustness of affect recognition and understanding. We will investigate nonlinear feature extraction schemes [4, 5] to identify biomarkers across complex biosignals (such as EEG, ECG and GSR recordings), as well as self-supervised learning methods [6] to model the interplay between physiological and contextual factors of affect-inducing tasks, such as music listening. This research has the potential to advance affective computing in areas such as health monitoring, human-computer interaction, and personalization systems [7, 8].

(Συνεργάτες: ΥΔ Χρήστος Γαρούφης, ΕΜΠ, Kleanthis Avramidis, USC, Los Angeles, και Δρ. Νάνσυ Ζλατίντση, ΕΜΠ)

References:

- [1] Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). [Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements](#). *Psychological science in the public interest*, 20(1), 1-68.
- [2] Soleymani, M., Pantic, M., & Pun, T. (2011). [Multimodal emotion recognition in response to videos](#). *IEEE transactions on affective computing*, 3(2), 211-223.
- [3] Dissanayake, V., Seneviratne, S., Rana, R., Wen, E., Kaluarachchi, T., & Nanayakkara, S. (2022). [Sigrep: Toward robust wearable emotion recognition with contrastive representation learning](#). *IEEE Access*, 10, 18105-18120.
- [4] Avramidis, K., Zlatintsi, A., Garoufis, C., & Maragos, P. (2021). [Multiscale fractal analysis on EEG signals for music-induced emotion recognition](#). In 2021 29th European Signal Processing Conference (EUSIPCO) (pp. 1316-1320).
- [5] Dimitriadis, D., Potamianos, A., & Maragos, P. (2009). [A comparison of the squared energy and Teager-Kaiser operators for short-term energy estimation in additive noise](#). *IEEE Transactions on signal processing*, 57(7), 2569-2581.
- [6] Avramidis, K., Garoufis, C., Zlatintsi, A., & Maragos, P. (2022). [Enhancing affective representations of music-induced EEG through multimodal supervision and latent domain adaptation](#). In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4588-4592).

- [7] Bone, D., Lee, C. C., Chaspari, T., Gibson, J., & Narayanan, S. (2017). [Signal processing and machine learning for mental health research and clinical applications \[perspectives\]](#). IEEE Signal Processing Magazine, 34(5), 196-195.
- [8] An, W. W., Shinn-Cunningham, B., Gamper, H., Emmanouilidou, D., Johnston, D., Jalobeanu, M., ... & Tashev, I. (2021). [Decoding music attention from "EEG headphones": A user-friendly auditory brain-computer interface](#). In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 985-989). IEEE.

Επικοινωνία Ανθρώπου-Ρομπότ και Αυτόνομη Πλοήγηση: Πολυτροπικές Ευφυείς Διεπαφές, Οραση, Ανάλυση Οπτικο-Ακουστικών Σημάτων, Ανίχνευση, Αναγνώριση, Πλοήγηση:

(*) R1. Αναγνώριση ανθρωπίνων δράσεων ή/και χειρονομιών για την αλληλεπίδραση ανθρώπου-ρομπότ, με εφαρμογές σε assistive robotics:

Μια από τις προκλήσεις για την ανάπτυξη ευφύων ρομπότ που αλληλεπιδρούν με φυσικό τρόπο με το περιβάλλον τους, είναι η ικανότητά τους να αναγνωρίζουν τις δράσεις και τις χειρονομίες που εκτελούν οι άνθρωποι ανά πάσα στιγμή. Στόχος της διπλωματικής αυτής είναι ο αυτόματος χρονικός εντοπισμός δράσεων/χειρονομιών σε πραγματικό χρόνο, η μοντελοποίηση και η αναγνώρισή τους με αξιοποίηση πολυτροπικών οπτικών-ακουστικών δεδομένων, όπως RGB-D ή/και ανθρώπινη πόζα. Οι μέθοδοι αυτές μπορούν να χρησιμοποιηθούν σε μεγάλες βάσεις δεδομένων ή/και σε ρομποτικές διατάξεις υποβοήθησης ηλικιωμένων και ατόμων με κινητικά προβλήματα.

(Συνεργάτες: ΥΔ Νίκος Κάρδαρης)

References:

- [1] I. Rodomagoulakis, N. Kardaris, V. Pitsikalis, E. Mavroudi, A. Katsamanis, A. Tsiami and P. Maragos, "[Multimodal Human Action Recognition in Assistive Human-Robot Interaction](#)", Proc. ICASSP-2016, Shanghai, China, Mar. 2016.
- [2] A. Zlatintsi, I. Rodomagoulakis, V. Pitsikalis, P. Koutras, N. Kardaris, X. Papageorgiou, C. Tzafestas and P. Maragos, "[Social Human-Robot Interaction for the Elderly: Two Real-life Use Cases](#)", Proc. ACM/IEEE Int'l Conf. on Human-Robot Interaction (HRI-2017), Vienna, Austria, March 2017.
- [4] J. Carreira, & A. Zisserman, "[Quo vadis, action recognition? a new model and the kinetics dataset](#)". Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017), Honolulu, Hawaii, July 2017
- [5] M. Xu, M. Gao, Y. Chen, L. S. Davis, D. J. Crandall, "[Temporal recurrent networks for online action detection](#)", Proc. IEEE Int'l Conference on Computer Vision (ICCV-2019), Seoul, Korea, Oct.-Nov. 2019.
- [6] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, Ahmed A. A. Osman, D. Tzionas, and M. J. Black, "[Expressive Body Capture: 3D Hands, Face, and Body from a Single Image](#)", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2019), Long Beach, CA, United States, June 2019.
- [7] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, Alex C. Kot, "[NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding](#)", in *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2019
- [8] G. Chalvatzaki, P. Koutras, A. Tsiami, C. S. Tzafestas and P. Maragos, "[i-Walk Intelligent Assessment System: Activity, Mobility, Intention, Communication](#)", Proc. 16th European Computer Vision Conference Workshops (ECCVW) – 8th Int'l Workshop on Assistive Computer Vision and Robotics (ACVR-2020), Aug. 2020.
- [9] V. Vasileiou, N. Kardaris, P. Maragos, [Exploring Temporal Context and Human Movement Dynamics for Online Action Detection in Videos](#), in Proc. 29th European Signal Processing Conference (EUSIPCO 2021), Dublin, Ireland, 2021.
- [10] G. Moustris, N. Kardaris, A. Tsiami, G. Chalvatzaki, P. Koutras, A. Dometios, P. Oikonomou, C. Tzafestas, P. Maragos, E. Efthimiou, X. Papageorgiou, S.-E. Fotinea, Y. Koumpouros, A. Vacalopoulou, E. Papageorgiou, A. Karavasili, F. Koureta, D. Dimou, A. Nikolakakis, K. Karaiskos and P. Mavridis, "[The i-Walk Lightweight Assistive Rollator: First Evaluation Study](#)", in *Frontiers in Robotics and AI*, vol. 8, pp. 272, Sep. 2021.

(*) R2. Αναγνώριση ανθρωπίνων δράσεων ή/και χειρονομιών για την αλληλεπίδραση ανθρώπου-ρομπότ σε πραγματικό χρόνο, με εφαρμογές σε social robotics

Στόχος είναι ο αυτόματος χρονικός εντοπισμός δράσεων/χειρονομιών (π.χ. gaze, handing, pointing), η μοντελοποίηση και η αναγνώρισή τους με αξιοποίηση πολυτροπικών οπτικών-ακουστικών δεδομένων (RGB-D, ανθρώπινη πόζα, audio/speech) . Οι παραπάνω πληροφορίες μπορούν να αξιοποιηθούν για την εξαγωγή χαρακτηριστικών high-level που συμβάλλουν στην πιο φυσική επικοινωνία ανθρώπου-ρομπότ, πχ εκτίμηση engagement κατά τη διάρκεια της αλληλεπίδρασης ή για την αυτόματη κατανόηση αλληλεπιδράσεων παιδιών με τις μητέρες τους. Εφαρμογές σε social robotics & child-robot interaction.

(Συνεργάτες: ΥΔ Δάφνη Αναγνωστοπούλου, ΥΔ Νίκη Ευθυμίου, ΥΔ Νίκος Κάρδαρης)

References:

- [1] A. Tsiami, P. Koutras, N. Efthymiou, P. P. Filntisis, G. Potamianos, and P. Maragos, “[Multi3: Multi-sensory Perception System for Multi-modal Child Interaction with Multiple Robots](#)”, *Proc. IEEE Int’l Conference on Robotics and Automation (ICRA-2018)*, Brisbane, Australia, May 2018.
- [2] N. Efthymiou, P.P Filntisis, P. Koutras, A. Tsiami, J. Hadfield, G. Potamianos, P. Maragos, “[ChildBot: Multi-Robot Perception and Interaction with Children](#)”, *Robotics and Autonomous Systems*, vol. 150, April 2022, 103975.
- [3] D. Anagnostopoulou, N. Efthymiou, C. Papaïliou and P. Maragos, “[Child Engagement Estimation in Heterogeneous Child-Robot Interactions Using Spatiotemporal Visual Cues](#)”, *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2022)*, Kyoto, Japan, October 2022.
- [4] N. Efthymiou, P. P. Filntisis, G. Potamianos, and P. Maragos, [Visual Robotic Perception System with Incremental Learning for Child–Robot Interaction Scenarios](#), *Technologies*, vol. 9, no. 4, article 86, Nov. 2021.
- [5] E. Marinoiu, M. Zafir, V. Olaru, and C. Sminchisescu. “[3D human sensing, action and emotion recognition in robot assisted therapy of children with autism](#)”, *Proc. CVPR 2018*.
- [6] S. Zahan, Z. Gilani, G. Mubashar Hassan, and A. Mian. “[Human Gesture and Gait Analysis for Autism Detection](#)”, *Proc. CVPRW*, 2023.

(* R3. Visual SLAM systems with geometric and/or semantic enhancements:

Τα συστήματα vSLAM [1,2] επιτυγχάνουν ταυτόχρονα την 3D ανακατασκευή του άγνωστου χώρου στον οποίο κινούνται, καθώς και την εκτίμηση της τροχιάς που διαγράφουν. Το vSLAM αποτελεί τη real-time εκδοχή του κλασικού προβλήματος της ΟΥ, Structure from Motion (SfM), ενώ έχει επίσης μελετηθεί παλιότερα υπό τη μορφή Time-of-Flight (ToF) μεθόδων (lasers ή sonars) και στο άλλο κλασικό πρόβλημα της Ρομποτικής, το SLAM. Πλέον, τα παραπάνω ενοποιούνται σε ένα σύστημα, με χρήση mono, stereo ή RGBD καμερών, καθώς και είναι δυνατό να συνδυαστούν με πληροφορία από LIDARs ή IMUs [7] οδηγώντας σε παραλλαγές συστημάτων vSLAM. Έμφαση δίνεται στην αξιοποίηση συνθετότερων γεωμετρικών δομών της εικόνας (ευθείες, επίπεδα, κωνικές) [3,4,6], σημείων φυγής για αποσύζευξη μεταφορικής και περιστροφικής συνιστώσας της πόζας [8] ή και σημασιολογικής πληροφορίας [4,5], με σκοπό την αύξηση της ευρωστίας, της ακρίβειας, του εύρους των περιβαλλόντων εφαρμογής, ή τη μείωση της υπολογιστικής πολυπλοκότητας. Ευρέως χρησιμοποιούμενα σύνολα δεδομένων σε εξωτερικούς χώρους είναι το [9] και σε εσωτερικούς τα [10,11].

(Συνεργάτες: ΥΔ Παναγιώτης Μέρμιγκας)

References:

- [1] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, “[ORB-SLAM: A Versatile and Accurate Monocular SLAM System](#)”, *IEEE Transactions on Robotics*, 2015.
- [2] D. Schlegel, M. Colosi and G. Grisetti, “[ProSLAM: Graph SLAM from a Programmer's Perspective](#)”, *IEEE International Conference on Robotics and Automation*, 2018.
- [3] F. Nardi, B. D. Corte and G. Grisetti, “[Unified Representation and Registration of Heterogeneous Sets of Geometric Primitives](#)”, *IEEE Robotics and Automation Letters*, 2019.
- [4] L. Nicholson, M. Milford and N. Sünderhauf, “[QuadricSLAM: Dual Quadrics From Object Detections as Landmarks in Object-Oriented SLAM](#)”, *IEEE Robotics and Automation Letters*, 2019.
- [5] I. Asmanis, P. Mermigkas, G. Chalvatzaki, J. Peters and P. Maragos, “[A Semantic Enhancement of Unified Geometric Representations for Improving Indoor Visual SLAM](#)”, *19th International Conference on Ubiquitous Robots (UR)*, 2022.
- [6] M. Hosseinzadeh, Y. Latif, T. Pham, N. Sünderhauf and I. Reid, “[Structure Aware SLAM Using Quadrics and Planes](#)”, *ACCV*, 2018.
- [7] A. Rosinol, M. Abate, Y. Chang and L. Carlone, “[Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping](#)”, *IEEE International Conference on Robotics and Automation*, 2020.
- [8] J. Liu and Z. Meng, “[Visual SLAM With Drift-Free Rotation Estimation in Manhattan World](#),” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6512-6519, Oct. 2020.
- [9] KIT, Visual Odometry/SLAM Evaluation, [Online]. Available: http://www.cvlibs.net/datasets/kitti/eval_odometry.php
- [10] TUM, RGB-D SLAM Dataset and Benchmark, [Online]. Available: <https://vision.in.tum.de/data/datasets/rgbd-dataset>
- [11] ICL-NUIM, Dataset for benchmarking VO and SLAM algorithms, [Online]. Available: <https://www.doc.ic.ac.uk/~ahanda/VaFRIC/iclnuim.html>

(* R4. Automatic 3D localization or annotation of electrical equipment in power plants using RGB images

Το θέμα αφορά την αυτόματη επισκόπηση ηλεκτρολογικού εξοπλισμού σε κέντρα υπερυψηλής τάσης με τη βοήθεια αυτόνομου ρομπότ. Πιο συγκεκριμένα, προτείνουμε δύο ξεχωριστά σενάρια εργασίας στα οποία θεωρούμε ότι η τρισδιάστατη αναπαράσταση του χώρου (π.χ. pointcloud που έχει παραχθεί με μεθόδους lidar SLAM) είναι ήδη γνωστή. Στο πρώτο σενάριο, επιθυμούμε την αυτόματη χωροθέτηση μιας άγνωστης έγχρωμης εικόνας και στη συνέχεια την online αντιστοίχιση rgb εικόνων με περιοχές ενός **ήδη επισημειωμένου** χάρτη (αυτόματη χωροθέτηση+αντιστοίχιση). Στο δεύτερο σενάριο, επιθυμούμε την εκ των προτέρων δημιουργία 3D μοντέλων των ηλεκτρολογικών στοιχείων μέσω της εκπαίδευσης κατάλληλων νευρωνικών δικτύων, τα οποία θα χρησιμοποιηθούν προκειμένου να “χρωματίσουν” περιοχές ενός **μη επισημειωμένου** χάρτη (αυτόματη επισημείωση).

(Συνεργάτες: ΥΔ Νίκος Κάρδαρης, Παναγιώτης Μέρμυγκας)

Οραση Υπολογιστών:

(* V1. Υλοποίηση συστήματος αναγνώρισης συμπεριφοράς (συναισθηματικής κατάστασης/χειρονομιών/δράσεων) και κατάτμηση εικόνων σε διαφορετικές σημασιολογικές περιοχές (π.χ. βλάστηση, δρόμοι, άνθρωποι) με χρήση βαθέων νευρωνικών δικτύων και μεθόδων domain adaptation και integration του συστήματος στο Robot-Operating-System software.

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ, και Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] P. P. Filntisis, N. Efthymiou, P. Koutras, G. Potamianos and P. Maragos, "[Fusing Body Posture With Facial Expressions for Joint Recognition of Affect in Child-Robot Interaction](#)," IEEE Robotics and Automation Letters, vol. 4, no. 4, pp. 4011-4018, Oct. 2019. DOI: [10.1109/LRA.2019.2930434](#) .
- [2] <https://www.ros.org/>
- [3] H. Gunes, & M. Pantic, “Automatic, dimensional and continuous emotion recognition”, ACM International Journal of Synthetic Emotions, 1(1), 68-99, 2010. <https://dl.acm.org/doi/10.4018/jse.2010101605> .
- [4] P. P. Filntisis, N. Efthymiou, G. Potamianos, P. Maragos, “[Emotion Understanding in Videos Through Body, Context, and Visual-Semantic Embedding Loss](#)”, Proc. ECCV Workshop, 2020.
- [5] I. Pikoulis, P. P. Filntisis, and P. Maragos. "[Leveraging Semantic Scene Characteristics and Multi-Stream Convolutional Architectures in a Contextual Approach for Video-Based Visual Emotion Recognition in the Wild.](#)" Proc. FG Conference, 2021
- [6] P. Antoniadis, I. Pikoulis, P. P. Filntisis, and P. Maragos. [An audiovisual and contextual approach for categorical and continuous emotion recognition in-the-wild](#). Proc. ICCV Workshop, 2021.
- [7] L. Wang et al. "Temporal segment networks: Towards good practices for deep action recognition." Proc ECCV 2016.
- [8] Kirillov, Alexander, et al. "Panoptic segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

V2. Three-dimensional Shape: Analysis, Modeling, Matching

References:

- [1] A. Bronstein, M. Bronstein, and R. Kimmel. [Numerical geometry of non-rigid shapes](#). Springer, 2008.
- [2] M. Breuss, A. Bruckstein and P. Maragos (Eds.), [Innovations for Shape Analysis: Models and Algorithms](#), Springer, 2013.
- [3] J. Sun, M. Ovsjanikov, L. Guibas, “[A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion](#)”, Eurographics Symposium on Geometry Processing 2009.

V3. Graph-theoretic Methods for Clustering and Segmentation

References:

- [1] C. G. Bampis, P. Maragos and A. C. Bovik, “[Graph-Driven Diffusion and Random Walk Schemes for Image Segmentation](#),” IEEE Transactions on Image Processing, vol. 26, no. 1, pp. 35-50, Jan. 2017.
- [2] C. Sakaridis, K. Drakopoulos and P. Maragos, “[Theoretical Analysis of Active Contours on Graphs](#),” SIAM J. Imaging Sciences, vol.10, no. 3, pp. 1475-1510, 2017.

V4. Sign-Language Understanding using Deep Learning

V4.1 Sign-Language Recognition using Deep Learning: Αναγνώριση χειρονομιών και συνεχών βίντεο νοηματικής γλώσσας με την χρήση βαθειών νευρωνικών δικτύων.

References:

- [1] S. Theodorakis, V. Pitsikalis and P. Maragos, “Dynamic–static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition”, *Image and Vision Computing*, vol.32, no.8, pp.533–549, Aug. 2014.
- [2] A. Roussos, S. Theodorakis, V. Pitsikalis and P. Maragos, “Dynamic Affine-Invariant Shape-Appearance Handshape Features and Classification in Sign Language Videos”, *Journal of Machine Learning Research*, vol. 14, pp. 1627-1663, June 2013.
- [3] O. Koller, N.C. Camgoz, H. Ney, and R. Bowden, “Weakly Supervised Learning with Multi-Stream CNN-LSTM-HMMs to Discover Sequential Parallelism in Sign Language Videos,” *IEEE Trans. PAMI*, 2020. DOI: 10.1109/TPAMI.2019.2911077
- [4] M. Parelli, K. Papadimitriou, G. Potamianos, G. Pavlakos and P. Maragos, “Exploiting 3D Hand Pose Estimation in Deep Learning-based Sign Language Recognition from RGB Videos”, *Proc. ECCV Workshop on Sign Language Recognition, Translation and Production (SLRTP 2020)*, Aug. 2020.
- [5] A. Kratimenos, G. Pavlakos and P. Maragos, “Independent Sign Language Recognition With 3d Body, Hands, And Face Reconstruction”, *ICASSP-2021*.
- [6] C. C. de Amorim, D. Macêdo and C. Zanchettin, “Spatial-Temporal Graph Convolutional Networks for Sign Language Recognition”, <https://arxiv.org/abs/1901.11164>, 2019.
- [7] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, <http://www.deeplearningbook.org>
- [8] *IEEE-EURASIP Summer School on Signal Processing Meets Deep Learning*, Capri, Italy, Sep. 2017.

(* V4.2 Sign-Language Generation using Deep Learning: Automatic generation of sign-language gestures from text and neural photorealistic rendering of videos.

The primary goal of this project is to revolutionize Sign Language Synthesis technology by developing an automatic system for transforming spoken language text into hyper-realistic sign language gestures. Current technology employs 3D avatars to translate spoken language into sign language, but it lacks realism in appearance, movement, and neglects the critical non-manual signals such as facial expressions. This project aims to address these limitations, generating highly realistic videos featuring not just hands, but also accurate body, head, and facial motions during signing. Furthermore, the system could render any real person, from celebrities to historical figures, as the virtual sign language interpreter. The successful implementation of this system could usher in a new era of sign language synthesis that significantly improves user engagement and experience, with potential applications in education, museums, and safety information. This ambitious yet feasible project builds on recent advances in neural rendering for facial and head re-enactment, along with state-of-the-art methods for deep 3D reconstruction of hands and full body. While there have been a few recent attempts at photo-realistic sign language synthesis, they suffer from limitations in modeling accuracy and realism, which this project seeks to overcome. The primary focus of this project is the automatic generation of sign language gestures from text.

(Συνεργάτες: Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ και Δρ. Αναστάσιος Ρούσσοις, Κύριος Ερευνητής, Ινστιτούτο Πληροφορικής, ΙΤΕ)

References:

- [1] Saunders, B., Camgoz, N. C., & Bowden, R. (2022). Signing at scale: Learning to co-articulate signs for large-scale photo-realistic sign language production. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5141-5151).
- [2] Forte, M. P., Kultits, P., Huang, C. H. P., Choutas, V., Tzionas, D., Kuchenbecker, K. J., & Black, M. J. (2023). Reconstructing Signing Avatars From Video Using Linguistic Priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12791-12801).
- [3] Tze, C. O., Filntisis, P. P., Dimou, A. L., Roussos, A., & Maragos, P. (2023). Neural Sign Reenactor: Deep Photorealistic Sign Language Retargeting. *CVPRW Workshop AI4CC(2023)*

V5. Shape from Shading. 3D Shape reconstruction from polarized images of unknown lighting and albedo:

References:

- [1] W. A.P. Smith, R. Ramamoorthi, S. Tozza, “Height-from-Polarisation with Unknown Lighting or Albedo”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.
- [2] S. Tozza, W. A.P. Smith, D. Zhu, R. Ramamoorthi, E. R. Hancock, “[Linear Differential Constraints for Photo-polarimetric Height Estimation](#)”, Proc. ICCV 2017.
- [3] S. Tozza, M. Falcone, “[Analysis and approximation of some Shape-from-Shading models for non-Lambertian surfaces](#)”, Journal of Mathematical Imaging and Vision, 55(2): 153-178, 2016.
- [4] R. Or – El. G. Rosman, A. Wetzler, R. Kimmel, A. Bruckstein, “[RGBD-Fusion: Real-Time High Precision Depth Recovery](#)”, Proc. CVPR 2015.

V6. 3D Computer Vision for Faces and Other Deformable Objects

(* V6.1 3D face modeling/reconstruction and applications (photorealistic avatars, talking heads)

The human face is one of the most commonly-considered objects in Computer Vision and Graphics. Modelling and reconstructing the detailed 3D shape and dynamics of the human face has numerous applications, such as augmented reality, performance capture, computer games, visual effects, human-computer interaction, computer-aided craniofacial surgery, rehabilitation and research in psychology, to name a few. During the last years we have proposed state-of-the-art methods for 3D face modelling, 3D face reconstruction from in-the-wild images and videos, facial expression recognition and photo-realistic synthesis of facial videos, see e.g. [1-5]. There can be various interesting projects related to the aforementioned state-of-the-art methods. These methods have yielded promising results, but there are many ways that they can be improved, extended, studied further or applied to different problems.

(Συνεργάτες: Δρ. Αναστάσιος Ρούσσος, Κύριος Ερευνητής, Ινστιτούτο Πληροφορικής, ΙΤΕ, και Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] M.C. Doukas, M.R. Koujan, V. Sharmanska, A. Roussos, and S. Zafeiriou, “[Head2Head++: Deep Facial Attributes Re-Targeting](#)”, IEEE Transactions on Biometrics, Behavior, and Identity Science (IEEE T-BIOM), Volume 3, Issue 1, pp 31-43, January 2021.
- [2] M.R. Koujan, ..., A. Roussos, “[Real-time Facial Expression Recognition “In The Wild” by Disentangling 3D Expression from Identity](#)”, Proc. Int’l Conf. Automatic Face and Gesture Recognition (FG 2020).
- [3] M.R. Koujan, A. Roussos and S. Zafeiriou. “[DeepFaceFlow: In-the-Wild Dense 3D Facial Motion Estimation](#)”, International Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [4] J. Booth, A. Roussos et al, “[3D reconstruction of in-the-Wild Faces in Images and Videos](#)”, PAMI 2018.
- [5] J. Booth, A. Roussos, ..., S. Zafeiriou, “[Large Scale 3D Morphable Face Models](#)”, IJCV 2017.
- [6] S. Ploumpis et al, “Towards a complete 3D morphable model of the human head”, <https://arxiv.org/abs/1911.08008> (CVPR and PAMI reprint).

(* V7. Style Adaptation from Real to Synthetic Images with Generative Adversarial Networks

Generative adversarial networks have been widely used to generate new, synthetic instances of data that can be used as real data in many problems and for many applications. SoftGrip is a European project that aims to facilitate delicate harvesting and boost production of mushrooms by developing a smart soft robotic gripper [1]. To accomplish that, the system must be able to estimate mushroom positions and poses. However, in order to develop such a system there is a great need for annotated mushroom data. These annotations would be tremendously time consuming and difficult to acquire. Therefore, the need for synthetic data is created. This thesis aims to leverage state of the art style transfer techniques [1,2,3] to produce high quality synthetic annotated data of mushrooms that can be used to train detection and pose estimation models. An existing simple pipeline to generate synthetic mushroom data will be extended and combined with real mushroom images to improve the resulting data quality.

(Συνεργάτες: Δρ. Γιώργος Πετσινάς, ΕΜΠ, ΥΔ Δάφνη Αναγνωστοπούλου, ΕΜΠ)

References:

- [1] Karras, Tero & Laine, Samuli & Aila, Timo. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. 4396-4405. 10.1109/CVPR.2019.00453.
- [2] Patashnik, Or & Wu, Zongze & Shechtman, Eli & Cohen-Or, Daniel & Lischinski, Dani. (2021). StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery.
- [3] P. Isola, J. -Y. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.
- [4] D. Anagnostopoulou, N. Efthymiou, C. Papailiou and P. Maragos, "Engagement Estimation During Child Robot Interaction Using Deep Convolutional Networks Focusing on ASD Children", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4th Agriculture Vision Workshop, 2023*
- [5] G Retsinas, N Efthymiou, P Maragos, 'Mushroom Segmentation and 3D Pose Estimation From Point Clouds Using Fully Convolutional Geometric Features and Implicit Pose Encoding', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4th Agriculture Vision Workshop, 2023*
- [6] <https://www.softgrip-project.eu/>

(* V8. 6DoF Pose Estimation of Objects with Emphasis on Dense Implicit Embeddings.

Ανάπτυξη αλγορίθμων εντοπισμού 3D πόζας αντικειμένων με τεχνικές βαθιάς μάθησης. Συγκεκριμένα ο εντοπισμός πόζας θα γίνεται είτε από (monocular) RGB εικόνες είτε από RGB-D εικόνες και θα δοθεί έμφαση στην κωδικοποίηση της πληροφορίας πόζας με ερμηνεύσιμες αναπαραστάσεις σε επίπεδο pixel ([3]). Ιδιαίτερη προσοχή θα δοθεί στον εντοπισμό πόζας συμμετρικών αντικειμένων.

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] Peng, Sida, et al. "Pvnet: Pixel-wise voting network for 6dof pose estimation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
- [2] He, Yisheng, et al. "Pvn3d: A deep point-wise 3d keypoints voting network for 6dof pose estimation." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
- [3] Retsinas, George, Niki Efthymiou, and Petros Maragos. "Mushroom Segmentation and 3D Pose Estimation From Point Clouds Using Fully Convolutional Geometric Features and Implicit Pose Encoding." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

(* V9. Image Manipulation via Guiding Attention Maps for Vision-Language Models.

Μελέτη συστημάτων κοινής περιγραφής πληροφορίας κειμένου/όρασης, όπως CLIP[1], και δημιουργίας εικόνων από κείμενο, π.χ. Imagen[2], για την εξαγωγή συμπερασμάτων (explainable-AI) για την μορφή των χαρτών προσοχής (attention masks). Επέκταση τεχνικών επεξεργασίας εικόνων (image editing) μέσω εντολών σε μορφή κειμένου (text prompts) που βασίζονται σε attention λογικές (π.χ. [3]).

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ, Δρ. Παναγιώτης Φιλντίσης, ΕΜΠ)

References:

- [1] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PMLR, 2021.
- [2] Saharia, Chitwan, et al. "Photorealistic text-to-image diffusion models with deep language understanding." *Advances in Neural Information Processing Systems* 35 (2022): 36479-36494.
- [3] Mokady, Ron, et al. "Null-text inversion for editing real images using guided diffusion models." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

Επεξεργασία Ακουστικών σημάτων (π.χ. μουσικής, φωνής) ή Οπτικών σημάτων σε Πολυμεσικά περιβάλλοντα:

AVM1. Υπολογιστική επεξεργασία μουσικών σημάτων και γενικότερα μουσικής πληροφορίας
(Συνεργάτες: Δρ. Νάνσυ Ζλατίντση και ΥΔ Χρήστος Γαρούφης, ΕΜΠ)

(*) AVM1.1: Αυτόματη αναγνώριση και εντοπισμός μουσικών χαρακτηριστικών.

Στόχος αυτής της διπλωματικής εργασίας είναι η εκπαίδευση νευρωνικών δικτύων που θα αναγνωρίζουν είτε το ηχητικό (πχ όργανο ή μουσικό είδος), είτε το σημασιολογικό (πχ συναίσθημα) περιεχόμενο ενός μουσικού κομματιού. Συγκεκριμένα, μπορεί να ακολουθηθεί είτε μία task-specific προσέγγιση, με εκπαίδευση σε ένα συγκεκριμένο σύνολο δεδομένων [1-4], είτε μία προσέγγιση αυτοεπιβλεπόμενης εκμάθησης αναπαραστάσεων, μέσω αντιθετικής μάθησης (self-supervised contrastive learning) [5-9]. Στην περίπτωση αυτή, μπορούν να χρησιμοποιηθούν δεδομένα τόσο από την ίδια (ακουστική) τροπικότητα, όσο και διαφορετικών τροπικοτήτων κατά τη διαδικασία εκπαίδευσης [10, 11], όπως για παράδειγμα των αποσπασμάτων μουσικής και των αντίστοιχων στίχων.

Datasets:

- MagnaTagATune: <https://mirg.city.ac.uk/codeapps/the-magnatagatune-dataset>
- MTG-Jamendo: <https://github.com/MTG/mtg-jamendo-dataset>
- DALI: <https://github.com/gabolsgabs/DALI>

References:

- [1] Choi, K., G. Fazekas, M. Sandler, and K. Cho. "[Convolutional Recurrent Neural Networks for Music Classification](#).", in Proc. ICASSP 2017
- [2] Kratimenos, A., K. Avramidis, C. Garoufis, A. Zlatintsi, and P. Maragos. "[Augmentation Methods On Monophonic Audio For Instrument Classification In Polyphonic Music](#)", in Proc. EUSIPCO–2020.
- [3] Lu, W. T., Wang, J. C., Won, M., Choi, K., & Song, X. "[SpecTNT: A time-frequency transformer for music audio](#)", in Proc. ISMIR 2021
- [4] Damböck, M., Vogl, R., & Knees, P. "[On the impact and interplay of input representations and network architectures for automatic music tagging](#)", in Proc. ISMIR 2022
- [5] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G.. "[A simple framework for contrastive learning of visual representations](#)", in Proc. ICML 2020
- [6] Saeed, A., Grangier, D., & Zeghidour, N. "[Contrastive learning of general-purpose audio representations](#)". In Proc. ICASSP 2021
- [7] Spijkervet, Janne, and John Ashley Burgoyne. "[Contrastive learning of musical representations](#).", in Proc. ISMIR 2021
- [8] Alonso-Jiménez, Pablo, Xavier Serra, and Dmitry Bogdanov. "[Music representation learning based on editorial metadata from discogs](#).", in Proc. ISMIR 2022
- [9] Garoufis, Christos, Athanasia Zlatintsi, and Petros Maragos. "[Multi-Source Contrastive Learning from Musical Audio](#).", in Proc. SMC 2023
- [10] Manco, Ilaria, Emmanouil Benetos, Elio QUINTON and Gyorgy Fazekas, "[Contrastive audio-language learning for music](#).", in Proc. ISMIR 2022
- [11] Avramidis, Kleanthis, Shanti Stewart, and Shrikanth Narayanan. "[On the Role of Visual Context in Enriching Music Representations](#).", in Proc ICASSP 2023

(*) AVM1.2: Ανάπτυξη συστήματος εξατομικευμένων μουσικών προτάσεων

Αντικείμενο αυτής της διπλωματικής εργασίας είναι η δημιουργία ενός συστήματος μουσικών προτάσεων (music recommendation) ή συνέχισης playlists (playlist continuation), βάσει είτε μουσικών χαρακτηριστικών ή μουσικής ομοιότητας [1-3], είτε μεθόδους collaborative filtering [4-5]. Για την ακουστική/μουσική ομοιότητα, πιθανές επιπρόσθετες κατευθύνσεις αποτελούν η χρήση sequence modelling [6] στα ακουστικά κομμάτια της κάθε playlist, καθώς και η διερεύνηση της μουσικής ομοιότητας μέσω των αναπαραστάσεων προεκπαιδευμένων μοντέλων σε άλλα tasks μουσικής ταξινόμησης [7,8]. Επιπρόσθετα, μία δυνατότητα αφορά την προσθήκη εξατομικευσης των προτάσεων, μέσω μεθοδολογιών ενεργής μάθησης (active learning) [9], ή

αποσύνδεσης (decoupling) των διαφόρων μουσικών παραμέτρων, όπως η χροιά, το είδος, ή το επαγόμενο συναίσθημα [3,10].

Datasets:

- Hit Song Prediction Dataset: <https://zenodo.org/record/3258042>
- MusAV: <https://mtg.github.io/musav-dataset>
- MPD: <https://research.atspotify.com/2020/09/the-million-playlist-dataset-remastered/>

References:

- [1] McFee, B., L. Barrington, and G. R. G. Lanckriet, "[Learning Content Similarity for Music Recommendation](#)", *IEEE Trans. on Audio, Speech & Language Processing* 20, no.8 (2012): 2207-2218
- [2] Huang, Q., A. Jansen, L. Zhang, D.P. Ellis, R.A. Saurous, and J. Anderson, "[Large-scale weakly-supervised content embeddings for music recommendation and tagging](#)". In *Proc ICASSP 2020*
- [3] Lee, J., Bryan, N. J., Salamon, J., Jin, Z., & Nam, J. . "[Disentangled multidimensional metric learning for music similarity](#)." In *Proc. ICASSP 2020*
- [4] W. Bendada, G. Salha-Galvan, T. Bouabça, and T. Cazenave, "[A Scalable Framework for Automatic Playlist Continuation on Music Streaming Services](#)". In *Proc. SIGIR 2023*
- [5] M. Volkovs, H. Rai, Z. Cheng, G. Wu, Y. Lu, and S. Sanner, "[Two-stage Model for Automatic Playlist Continuation at Scale](#)" In *Proc.of the ACM Recommender Systems Challenge 2018, 2018.*
- [6] Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. [wav2vec 2.0: A framework for self-supervised learning of speech representations](#), in *Proc. NeurIPS 2020*
- [7] Spijkervet, Janne, and John Ashley Burgoyne. "[Contrastive learning of musical representations](#).", in *Proc. ISMIR 2021*
- [8] Manco, Iaria, Emmanouil Benetos, Elio Quintion and Gyorgy Fazekas, "[Contrastive audio-language learning for music](#).", in *Proc. ISMIR 2022*
- [9] Cañón, J. S. G., Cano, E., Yang, Y. H., Herrera, P., & Gómez, E. . [Let's agree to disagree: Consensus Entropy Active Learning for Personalized Music Emotion Recognition](#), in *Proc. ISMIR 2021*
- [10] Wang, X., Chen, H., Zhou, Y., Ma, J., & Zhu, W. (2022). [Disentangled representation learning for recommendation](#). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 408-424.

AVM1.3: Διαχωρισμός μουσικών πηγών

Αντικείμενο αυτής της διπλωματικής εργασίας αποτελεί το πρόβλημα του διαχωρισμού μουσικών πηγών (music source separation), της απομόνωσης δηλαδή των φωνητικών, ή των επιμέρους οργάνων, που αποτελούν ένα μουσικό κομμάτι. Το τελευταίο χρονικό διάστημα παρατηρείται μία τάση αντικατάστασης, ή ενίσχυσης, των παραδοσιακών, state-of-the-art αρχιτεκτονικών [1-5] με καθαρά παραγωγικές (generative) προσεγγίσεις, όπως αυτή των diffusion models [6-8]. Μία άλλη ενεργή πιθανή κατεύθυνση του προβλήματος είναι αυτή του query-based διαχωρισμού πηγών, όπου στόχος είναι ο διαχωρισμός μίας πηγής-στόχου από ένα ηχητικό μίγμα [9-10], ενώ ενδιαφέρον παρουσιάζει και η χρήση κατάλληλων μεθοδολογιών, εμπνευσμένων από την επεξεργασία ψηφιακού σήματος, με στόχο τη μείωση του υπολογιστικού κόστους (αριθμός παραμέτρων, ή χρόνος εκτέλεσης) των μοντέλων [11-12].

Datasets:

- musdb18: <https://sigsep.github.io/datasets/musdb.html>
- URMP: <https://labsites.rochester.edu/air/projects/URMP.html>
- EnsembleSet: <https://zenodo.org/record/6519024>

References:

- [1] Stoller, D., S. Ewert, and S. Dixon. "[Wave-U-Net: A Multi-Scale Neural Network for End-to-End Audio Source Separation](#)." In *Int'l Soc. for Music Inform. Retrieval Conf. (ISMIR)*, Paris, France, 2018.
- [2] Kong, Q., Cao, Y., Liu, H., Choi, K., & Wang, Y. [Decoupling magnitude and phase estimation with deep resunet for music source separation](#), in *Proc. ISMIR 2021*
- [3] Défossez, A.. [Hybrid spectrogram and waveform source separation](#), in *Proc. MDX Workshop 2021*
- [4] Papantonakis, P., Garoufis, C., & Maragos, P. [Multi-band Masking for Waveform-based Singing Voice Separation](#). In *Proc. EUSIPCO 2022*

- [5] Luo, Yi, and Jianwei Yu. "[Music Source Separation With Band-Split RNN.](#)" *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2023).
- [6] Plaja-Roglans, G., Miron, M., & Serra, X. [A diffusion-inspired training strategy for singing voice extraction in the waveform domain.](#), in Proc. ISMIR 2022
- [7] Mariani, G., Tallini, I., Postolache, E., Mancusi, M., Cosmo, L., & Rodolà, E. [Multi-source diffusion models for simultaneous music generation and separation.](#) *arXiv preprint arXiv:2302.02257.*
- [8] Lutati, S., Nachmani, E., & Wolf, L. [Separate and diffuse: Using a pretrained diffusion model for improving source separation.](#) *arXiv preprint arXiv:2301.10752.*
- [9] Lin, L., Kong, Q., Jiang, J., & Xia, G. [A unified model for zero-shot music source separation, transcription and synthesis.](#), in Proc. ISMIR 2021
- [10] Kong, Q., Chen, K., Liu, H., Du, X., Berg-Kirkpatrick, T., Dubnov, S., & Plumbley, M. D.,. [Universal Source Separation with Weakly Labelled Data.](#) *arXiv preprint arXiv:2305.07447.*
- [11] Ditter, David, and Timo Gerkmann. "[A multi-phase gammatone filterbank for speech separation via TASNet.](#)", in Proc. ICASSP 2020
- [12] Mathieu, F., Courtat, T., Richard, G., & Peeters, G. [Phase shifted bedrosian filterbank: An interpretable audio front-end for time-domain audio source separation.](#), in *Proc. ICASSP 2022*

AVM2. Audio-Visual Multi-tasking Network: Action Recognition, Saliency Estimation and Video Summarization: Οπτικοακουστικό δίκτυο που πραγματοποιεί ταυτόχρονα αναγνώριση δράσεων, εκτίμηση του saliency και περίληψη

References:

- [1] P. Koutras and P. Maragos, "[SUSiNet: See, Understand and Summarize It](#)", *Proc. CVPR Workshop* 2019.
- [2] A. Tsiami, P. Koutras and P. Maragos, "[STAViS: Spatio-Temporal AudioVisual Saliency Network](#)", *Proc. CVPR 2020.*
- [3] G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas and Y. Avrithis, [Multimodal Saliency and Fusion for Movie Summarization based on Aural, Visual, and Textual Attention](#), *IEEE Trans. Multimedia*, Nov. 2013.
- [4] A. Zlatintsi, P. Koutras, G. Evangelopoulos, N. Malandrakis, N. Efthymiou, K. Pastra, A. Potamianos and P. Maragos, "[COGNIMUSE: a multimodal video database annotated with saliency, events, semantics and emotion with application to summarization](#)", *EURASIP Journal on Image and Video Processing* (2017) 2017:54.

AVM3. Υπολογιστική επεξεργασία φωνητικών ή μουσικών ή βιοϊατρικών σημάτων με μη-γραμμικά μοντέλα AM-FM και αλγορίθμους πολυζωνικής ενεργειακής αποδιαμόρφωσης, και σχετικά μη-γραμμικά συστήματα.

References:

- [1] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "[Energy Separation in Signal Modulations with Application to Speech Analysis](#)", *IEEE Trans. Signal Processing*, vol.41, no.10, pp.3024-3051, Oct. 1993.
- [2] A. Potamianos and P. Maragos, "[Speech Processing Applications Using an AM-FM Modulation Model](#)", *Speech Communication*, vol.28, no.3, pp.195-209, July 1999.
- [3] D. Dimitriadis, P. Maragos, and A. Potamianos, "[Robust AM-FM Features for Speech Recognition](#)", *IEEE Signal Processing Letters*, vol.12, no.9, pp.621-624, Sep. 2005.
- [4] J. H.L. Hansen and S. Patil, [Speech Under Stress: Analysis, Modeling and Recognition](#), Springer LNAI 4343, 2007.
- [5] D. Dimitriadis, P. Maragos and A. Potamianos, "[On the Effects of Filterbank Design and Energy Computation on Robust Speech Recognition](#)", *IEEE Trans. Audio, Speech and Language Processing*, vol.19, pp.1504-1516, Aug.2011.
- [6] A. Zlatintsi and P. Maragos, "[Comparison of Different Representations Based on Nonlinear Features for Music Genre Classification](#)", *Proc. EUSIPCO-2014*, Lisbon, Portugal, Sep. 2014.
- [7] T. Chaspari, D. Dimitriadis and P. Maragos, "[Emotion Classification of Speech Using Modulation Features](#)", *Proc. EUSIPCO-2014*, Lisbon, Portugal, Sep. 2014.
- [8] I. Rodomagoulakis and P. Maragos, "[On the Improvement of Modulation Features Using Multi-Microphone Energy Tracking for Robust Distant Speech Recognition](#)", *Proc. EUSIPCO-2017*, Kos, Greece, Aug. 2017.
- [9] H. B. Sailor, M. R. Kamble and H. A. Patil, "[Auditory Filterbank Learning for Temporal Modulation Features in Replay Spoof Speech Detection](#)", in *Proc. Interspeech 2018.*

AVM4. Υπολογιστική επεξεργασία φωνητικών ή μουσικών ή βιοϊατρικών σημάτων με μη-γραμμικά μοντέλα από φράκταλς και χάος, και σχετικά μη-γραμμικά συστήματα.

References:

- [1] P. Maragos and A. Potamianos, "[Fractal Dimensions of Speech Sounds: Computation and Application to Automatic Speech Recognition](#)", *Journal of Acoustical Society of America*, vol.105 (3), pp.1925--1932, March 1999.
- [2] V. Pitsikalis and P. Maragos, "[Filtered Dynamics and Fractal Dimensions for Noisy Speech Recognition](#)", *IEEE Signal Processing Letters*, vol.13, no.11, pp.711-714, Nov. 2006..
- [3] V. Pitsikalis and P. Maragos, "[Analysis and Classification of Speech Signals by Generalized Fractal Dimension Features](#)", *Speech Communication*, vol.51, no.12, pp.1206-1223, Dec. 2009.
- [4] A. Zlatintsi and P. Maragos, "[Multiscale Fractal Analysis of Musical Instrument Signals with Application to Recognition](#)", *IEEE Trans. Audio, Speech and Language Processing*, vol.21, no.4, pp.737-748, Apr. 2013.
- [5] Karmele López-de-Ipina et al., "[Feature selection for spontaneous speech analysis to aid in Alzheimer's disease diagnosis: A fractal dimension approach](#)", *Computer Speech and Language* 30, pp.43–60, 2015.
- [6] E. Tzinis, G. Paraskevopoulos, C. Baziotis, A. Potamianos, "[Integrating Recurrence Dynamics for Speech Emotion Recognition](#)", in *Proc. Interspeech 2018*.
- [7] K. Avramidis, A. Zlatintsi, C. Garoufis, and P. Maragos, "[Multiscale Fractal Analysis on EEG Signals for Music-Induced Emotion Recognition](#)", *Proc. EUSIPCO, 2021*.
- [8] W. Gilpin, "Deep Reconstruction of Strange Attractors from Time Series", NIPS 2020, [arXiv:2002.05909 \(2020\)](#).

Συστήματα, Δίκτυα, Γράφοι, Optimization & Learning:

S1. Μη-γραμμικά δυναμικά συστήματα που χρησιμοποιούν max-plus άλγεβρα και finite-state automata με εφαρμογές σε ένα πρόβλημα από τις περιοχές ανίχνευσης, βελτιστοποίησης, δικτύων, ελέγχου, θεωρία γράφων, κ.ά.

References:

- [1] R. Cuninghame-Green, [Minimax Algebra](#), Springer-Verlag, New York, 1979.
- [2] P. Butkovič. [Max-linear Systems: Theory and Algorithms](#). Springer, 2010.
- [3] P. Maragos, "[Dynamical Systems on Weighted Lattices: General Theory](#)", *Mathematics of Control, Signals and Systems*, 29:21, 2017.
- [4] I. Kordonis, P. Maragos, G. P. Papavassilopoulos, "[Stochastic Stability in Max-Product and Max-Plus Systems with Markovian Jumps](#)," *Automatica* 92 (2018) 123–132.

S2. Sparsity and Max-plus Algebra, Statistical Analysis of Tropical Estimators

(Συνεργάτες: Νίκος Τσιλιβής, Courant Institute of New York)

- [1] A. Tsiamis and P. Maragos, "[Sparsity in Max-plus Algebra](#)", *Discrete Events Dynamic Systems*, 29:163–189, May 2019. <https://doi.org/10.1007/s10626-019-00281-1>
- [2] N. Tsilivis, A. Tsiamis and P. Maragos, "[Sparsity in Max-Plus Algebra and Applications in Multivariate Convex Regression](#)", *Proc. ICASSP-2021*, June 2021.
- [3] N. Tsilivis, A. Tsiamis and P. Maragos, "Sparse Approximate Solutions to Max-Plus Equations", In: J. Lindblad et al. (Eds), *Discrete Geometry and Mathematical Morphology: DGMM-2021*, LNCS 12708, Springer, Cham, 2021, pp. 538-550. https://doi.org/10.1007/978-3-030-76657-3_39

S3. Signal Processing on Graphs

References:

- [1] A. Sandryhaila and J. Moura, "Discrete Signal Processing on Graphs", *IEEE Trans. Signal Processing*, Apr. 2013.
- [2] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The Emerging Field of Signal Processing on Graphs", *IEEE Signal Processing Magazine*, May 2013.
- [3] G. Giannakis, Y. Shen and G. Karanikolas, "Topology Identification and Learning Over Graphs: Accounting for Nonlinearities and Dynamics", *Proc. IEEE*, vol.106, May 2018.

Machine Learning:

L1. Machine Learning, Deep Neural Nets, and Geometry

(*) L1.1 Geometry, Pruning, Training, Applications of DNNs with PWL Activations and/or Morphological Networks for Classification or Regression Problems

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ, Μάνος Θεοδόσης, Harvard University)

- Τα τροπικά μαθηματικά μπορούν να δώσουν κλειστές λύσεις για piecewise-linear approximations [4]. Μετέπειτα εργασίες [9] επεκτείνουν το framework ώστε να γίνεται sparse approximation των πολυδιάστατων επιφανειών. Όμως, η επιλογή των στοιχείων για το sparse approximation γίνεται από μια δοσμένη “βάση” πιθανών στοιχείων. Στο προκείμενο θέμα εξετάζουμε τρόπους να μάθουμε από τα δεδομένα τη βάση των πιθανών στοιχείων για το sparse approximation.
- Ενώ τα τροπικά μαθηματικά μπορούν να χρησιμοποιηθούν για piecewise-linear approximations [4], τα approximations είναι πάντα convex. Κάτι τέτοιο περιορίζει την ευρεία εφαρμογή των ανεπτυγμένων μεθόδων, και παράλληλα δεν επιτρέπει το approximation με αυθαίρετη ευκρίνεια. Σε αυτό το θέμα εξετάζουμε τρόπους για να αρθεί αυτός ο περιορισμός (π.χ. difference of convex functions) ή τρόπους να χαρακτηριστεί το approximation error έναντι κάποιων βέλτιστων μεθόδων (π.χ. με upper και lower envelopes).
- Το pruning μεμονωμένων βαρών ενός νευρωνικού δικτύου, ενώ μειώνει το μέγεθος του δικτύου, δεν οδηγεί σε speedups κατά τη διάρκεια του inference [11]. Παράλληλα, pruning μέθοδοι που αντιμετωπίζουν το πρόβλημα globally συνήθως αγνοούν την εσωτερική δομή του δικτύου, και μέθοδοι που επιτρέπουν learnable sparsity για τα δίκτυα [12] οδηγούν σε καλύτερα αποτελέσματα. Έχοντας ως έναυσμα τα παραπάνω, στο προκείμενο θέμα διπλωματικής εξετάζουμε το pruning νευρωνικών δικτύων χρησιμοποιώντας learnable structured sparsity.

References:

- [1] V. Charisopoulos and P. Maragos, “[Morphological Perceptrons: Geometry and Training Algorithms](#),” *Proc. ISMM 2017*. Lecture Notes in Computer Science, vol. 10225, Springer, 2017.
- [2] V. Charisopoulos and P. Maragos, “[A tropical approach to neural networks with piecewise linear activations](#)”, arXiv:1805.08749, 2018.
- [3] E. Theodosis and P. Maragos, “[Analysis of the Viterbi Algorithm Using Tropical Algebra and Geometry](#)”, Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC-18), Greece, June 2018.
- [4] P. Maragos and E. Theodosis, “[Multivariate Tropical Regression and Piecewise-Linear Surface Fitting](#)”, *Proc. 45th IEEE Int’l Conf. on Acoustics, Speech and Signal Processing (ICASSP-2020)*, May 2020.
- [5] G. Smyrnis, G. Retsinas, P. Maragos, “[Maxpolynomial Division with Application to Neural Network Simplification](#)”, Proc. IEEE ICASSP, May 2020.
- [6] G. Smyrnis and P. Maragos, “[Multiclass Neural Network Minimization Via Tropical Newton Polytope Approximation](#)”, *Proc. 37th Int’l Conf. on Machine Learning (ICML-2020)*, July 2020.
- [7] P. Maragos and E. Theodosis, “[Tropical Geometry and Piecewise-Linear Approximation of Curves and Surfaces on Weighted Lattices](#)”, in *Shape Analysis: Euclidean, Discrete and Algebraic Geometric Methods*, edited by M. Breuss, A. Bruckstein, C. Kiselman and P. Maragos, Springer, to appear.
- [8] N. Dimitriadis and P. Maragos, “[Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints](#)”, *Proc. ICASSP-2021*, June 2021.
- [9] N. Tsilivis, A. Tsiamis and P. Maragos, “[Sparsity in Max-Plus Algebra and Applications in Multivariate Convex Regression](#)”, *Proc. ICASSP-2021*, June 2021.
- [10] N. Tsilivis, A. Tsiamis and P. Maragos, “Sparse Approximate Solutions to Max-Plus Equations”, In: J. Lindblad et al. (Eds), *Discrete Geometry and Mathematical Morphology: DGMM-2021*, LNCS 12708, Springer, Cham, 2021, pp. 538-550. https://doi.org/10.1007/978-3-030-76657-3_39
- [11] W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li, “[Learning Structured Sparsity in Deep Neural Networks](#)”, *Advances of Neural Information Processing Systems*, 2016.
- [12] A. Kusupati, V. Ramanujan, R. Somani, M. Wortsman, P. Jain, S. Kakade, A. Farhadi, “[Soft Threshold Weight Reparameterization for Learnable Sparsity](#)”, *International Conference on Machine Learning*, 2020.
- [13] P. Maragos, V. Charisopoulos, E. Theodosis, “Tropical Geometry and Machine Learning”, *Proceedings of the IEEE*, 109 (5), pp. 728-755, 2021. <https://ieeexplore.ieee.org/document/9394420>

(* L1.2 Tropical Properties of Deep Neural Networks

- Μελέτη τροπικών ιδιοτήτων νευρωνικών δικτύων, με έμφαση σε βαθιές αρχιτεκτονικές, CNNs και ResNets.
- Χρήση τροπικών ιδιοτήτων για μελέτη συμπεριφοράς δικτύων σε προβλήματα μηχανικής μάθησης.
- Πρακτική εφαρμογή στα ανωτέρω προβλήματα (ενδεικτικό πρόβλημα: συμπίεση νευρωνικών δικτύων).

(Συνεργάτες: Δρ. Ιωάννης Κορδώνης, ΕΜΠ, Γιώργος Σμυρνής, University of Texas at Austin)

References:

- [1] V. Charisopoulos and P. Maragos, "[A tropical approach to neural networks with piecewise linear activations](#)", arXiv:1805.08749, 2018.
- [2] N. Dimitriadis and P. Maragos, "[Advances in Morphological Neural Networks: Training, Pruning and Enforcing Shape Constraints](#)", *Proc. IEEE ICASSP 2021*, June 2021.
- [3] P. Misiakos, G. Smyrnis, G. Retsinas and P. Maragos, "[Neural Network Approximation based on Hausdorff distance of Tropical Zonotopes](#)", *Proc. ICLR 2022*, April 2022.
- [4] G. Smyrnis, G. Retsinas, P. Maragos, "[Maxpolynomial Division with Application to Neural Network Simplification](#)", *Proc. IEEE ICASSP 2020*, May 2020.
- [5] G. Smyrnis and P. Maragos, "[Multiclass Neural Network Minimization Via Tropical Newton Polytope Approximation](#)", *Proc. 37th Int'l Conf. on Machine Learning (ICML-2020)*, July 2020.
- [6] L. Zhang, G. Naitzat, and L.-H. Lim. "[Tropical geometry of deep neural networks](#)", *Proc. 35th Int'l Conf. on Machine Learning (ICML-2018)*, July 2018.
- [7] P. Maragos, V. Charisopoulos, E. Theodosis, "Tropical Geometry and Machine Learning", *Proceedings of the IEEE*, 109 (5), pp. 728-755, 2021. <https://ieeexplore.ieee.org/document/9394420>

(* L2. Deep Neural Network Compression.

Ελαχιστοποίηση χωρικών/χρονικών απαιτήσεων σύγχρονων νευρωνικών δικτύων: Μελέτη του προβλήματος συμπίεσης νευρωνικών μέσω magnitude pruning των βαρών ενός NN. Έμφαση σε συναρτήσεις καταφλοιοποίησης και στην ευστάθεια της pruning μάσκας για μεγάλα ποσοστά συμπίεσης και επέκταση σε group sparsity λογικές. Ενδεικτικές εναλλακτικές κατευθύνσεις: quantization, weight sharing.

Αξιολόγηση σε δημοφιλείς βάσεις ταξινόμησης εικόνων, όπως:

<https://paperswithcode.com/sota/image-classification-on-cifar-100>.

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] G. Retsinas, A. Elafrou, G. Goumas and P. Maragos, "[RecNets: Depthwise-Recurrent Convolutional Neural Networks](#)", *BMVC*, 2019.
- [2] G. Smyrnis, G. Retsinas, P. Maragos, "[Maxpolynomial Division with Application to Neural Network Simplification](#)", *ICASSP*, 2020.
- [3] G. Retsinas, A. Elafrou, G. Goumas and P. Maragos. "[Weight Pruning via Adaptive Sparsity Loss](#)", arXiv preprint arXiv:2006.02768, 2020.
- [4] Y. Cheng, D. Wang, P. Zhou and T. Zhang, "[A survey of model compression and acceleration for deep neural networks](#)", arXiv preprint arXiv:1710.09282, 2017.
- [5] M. Rastegari, V. Ordonez, J. Redmon and A. Farhadi, "[Xnor-net: Imagenet classification using binary convolutional neural networks](#)", *ECCV*, 2016.

L3. Self-Supervision and contrastive learning.

Στο πλαίσιο αυτής της εργασίας θα γίνει μελέτη τεχνικών αυτο-επιβλεπόμενης μάθησης (self-supervised learning), για την εκμάθηση αναπαραστάσεων δεδομένων εισόδου προς χρήση σε πολλαπλά tasks. Ενδεικτική κατεύθυνση είναι η ανάπτυξη τεχνικών στον κλάδο της μάθησης με αντιδιαστολή (contrastive learning), η οποία είναι ευρέως χρησιμοποιούμενη για την εκπαίδευση state of the art βαθιών νευρωνικών δικτύων.

References:

- [1] T. Chen, S. Kornblith, M. Norouzi and G. Hinton, "[A simple framework for contrastive learning of visual representations](#)." *Proc. ICML*, 2020
- [2] J. D. Lee, Q. Lei, N. Saunshi and J. Zhuo, "[Predicting what you already know helps: Provable self-supervised learning](#)" *Proc. NeurIPS*, 2021.

- [3] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark and G. Krueger, "[Learning transferable visual models from natural language supervision](#)" *Proc. ICML*, 2021
- [4] T. Wang and P. Isola, "[Understanding contrastive representation learning through alignment and uniformity on the hypersphere](#)" *Proc. ICML*, 2020

(* L4. Deep hyperspectral unmixing. (Θ. Ροντογιάννης)

Η πολυδιάστατη απεικόνιση με τη χρήση υπερφασματικών αισθητήρων (hyperspectral sensors) και η επεξεργασία και ανάλυση των υπερφασματικών δεδομένων που προκύπτουν συναντάται σε ένα ευρύτατο φάσμα εφαρμογών τα τελευταία χρόνια. Ένα ενδιαφέρον πρόβλημα στο πεδίο αυτό είναι αυτό του φασματικού διαχωρισμού υπερφασματικών εικόνων (hyperspectral unmixing), που αποτελεί ουσιαστικά ένα είδος προβλήματος τυφλού διαχωρισμού πηγών (blind source separation). Ο φασματικός διαχωρισμός αποσκοπεί τόσο στην αναγνώριση, με βάση τις φασματικές τους υπογραφές, των αντικειμένων ή υλικών που εμφανίζονται στην περιοχή απεικόνισης, όσο και στην περιγραφή της κατανομής των υλικών πάνω στην εικόνα. Παραδοσιακά το πρόβλημα αυτό προσεγγίζεται με κατάλληλη μαθηματική μοντελοποίηση του μηχανισμού παραγωγής των υπερφασματικών δεδομένων και στην ανάπτυξη τεχνικών εκτίμησης των παραμέτρων του μοντέλου γέννησης των δεδομένων (model-based approach). Αξίζει να σημειωθεί ότι σημαντικές βελτιώσεις στις κλασσικές τεχνικές έχουν επιτευχθεί με την αξιοποίηση φυσικών περιορισμών που ικανοποιούν τα δεδομένα, όπως μη αρνητικότητα (nonnegativity), αραιότητα (sparsity), χαμηλή τάξη (low-rank), κλπ. Πρόσφατα, η χρησιμοποίηση τεχνικών βαθιάς μάθησης έχει δώσει νέα ώθηση στο πεδίο του φασματικού διαχωρισμού υπερφασματικών εικόνων. Στο πλαίσιο της εργασίας αυτής, βασικός στόχος είναι η ανάπτυξη νέων αποδοτικών τεχνικών φασματικού διαχωρισμού, συνδυάζοντας την εμπειρία που έχει αποκτηθεί σε κλασσικές μεθόδους με τις νέες ιδέες εξαγωγής γνώσης από τα δεδομένα που βασίζονται σε αρχιτεκτονικές και αλγόριθμους βαθιάς μάθησης.

References:

- [1] K.E. Themelis, A.A. Rontogiannis, K.D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semi-supervised hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 585-599, Feb. 2012.
- [2] P.V. Giampouras, K.E. Themelis, A.A. Rontogiannis, K.D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, issue 8, pp. 4775-4789, Aug. 2016.
- [3] C. Tsinos, A.A. Rontogiannis, K. Berberidis, "Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained nonnegative matrix factorization," *IEEE Transactions on Computational Imaging*, vol. 3, issue 2, pp. 160-174, June 2017.
- [4] X. Zhang, Y. Sun, J. Zhang, P. Wu, and L. Jiao, "Hyperspectral unmixing via deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 11, pp. 1755-1759, 2018.
- [5] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1427-1431, 2018.
- [6] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Deep generative endmember modeling: An application to unsupervised spectral unmixing," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 374-384, 2020.
- [7] B. Rasti, B. Koirala, P. Scheunders, and J. Chanussot, "Misticnet: Minimum simplex convolutional network for deep hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022.

(* L5. Representation Learning of Visual Entities and Generalization using Reinforcement Learning.

Εξαγωγή οπτικών χαρακτηριστικών (π.χ. σχήμα, χρώμα) απλών αντικειμένων (π.χ. σφαίρες, κύβοι) μέσω Reinforcement Learning σε περιβάλλοντα προσομοίωσης όπου ένας ρομποτικός βραχίονας αλληλεπιδρά με τα αντικείμενα αυτά, χωρίς καμία πρότερη γνώση των αντικειμένων. Επί της ουσίας θα «χτιστεί» ένα σύστημα όρασης από δεδομένα προσομοίωσης και θα γίνει μελέτη ικανότητας γενίκευσης εξαγόμενων χαρακτηριστικών.

(Συνεργάτες: Δρ. Γιώργος Πετσινάς, ΕΜΠ)

References:

- [1] Heuillet, Alexandre, Fabien Couthouis, and Natalia Díaz-Rodríguez. "Explainability in deep reinforcement learning." *Knowledge-Based Systems* 214 (2021): 106685.
- [2] Gershman, Samuel, Jonathan Cohen, and Yael Niv. "Learning to selectively attend." *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 32. No. 32. 2010.
- [3] Le Goff, Leni K., et al. "Building an affordances map with interactive perception." *Frontiers in Neurorobotics* 16 (2022): 504459.

(*) L6. Προβλήματα Προσέγγισης στην Τροπική Άλγεβρα

Η τροπική άλγεβρα και γεωμετρία μπορούν να χρησιμοποιηθεί για την αναπαράσταση σε συμπαγή μορφή και την ανάλυση πολλών νευρωνικών δικτύων (με κατά τμήματα γραμμική ενεργοποίηση π.χ. ReLU, Maxout, κλπ), καθώς και την περιγραφή συνήθων αλγορίθμων όπως ο Viterbi [1]. Επιπλέον, στην τροπική άλγεβρα, ορισμένα προβλήματα μπορούν να λυθούν πολύ αποδοτικά (π.χ. η λύση γραμμικού συστήματος είναι πολύ πιο σύντομη στην τροπική άλγεβρα από ότι στη συνήθη [2]). Η διπλωματική αυτή ασχολείται με προβλήματα προσέγγισης στην Τροπική Άλγεβρα, δηλαδή περιπτώσεις στις οποίες θέλουμε να προσεγγίσουμε μια τροπική απεικόνιση από δεδομένα. Σε αυτό περιλαμβάνονται ενδεικτικά ένα ή περισσότερα από τα παρακάτω:

- **Προβλήματα κυρτής/τροπικής παλινδρόμησης με outliers.** Υπάρχει αρκετή προηγούμενη δουλειά σε κυρτή/τροπική παλινδρόμηση ([3],[4],[5]). Στα γραμμικά μοντέλα υπάρχουν πολύ επιτυχημένοι αλγόριθμοι για γραμμική παλινδρόμηση με outliers (π.χ. iteratively reweighted least squares, RANSAC). Στόχος εδώ είναι να επεκταθούν/τροποποιηθούν οι αλγόριθμοι της γραμμικής παλινδρόμησης στην περίπτωση της κυρτής/τροπικής.

- **Παραγοντοποίηση/Προσέγγιση Πινάκων στην Τροπική Άλγεβρα.** Η προσέγγιση χαμηλής τάξης των πινάκων καθώς και η παραγοντοποίηση με πίνακες με μη αρνητικούς όρους (Nonnegative Matrix Factorization) έχουν βρει πολλές εφαρμογές, όπως για παράδειγμα η ανάλυση κειμένων, η συμπίεση και τα recommender systems. Έχει ενδιαφέρον να επεκταθεί η προσεγγιστική παραγοντοποίηση στην τροπική περίπτωση. Κάποιες προκαταρκτικές εργασίες είναι οι [6],[7],[8]. Στόχος της διπλωματικής θα είναι να προταθούν πιο αποτελεσματικοί αλγόριθμοι, και να επεκταθεί το πεδίο εφαρμογής σε περιπτώσεις που οι πίνακες που αναζητούμε έχουν επιπλέον ιδιότητες (π.χ. χαμηλή τάξη).

(Συνεργάτες: Δρ. Ιωάννης Κορδώνης, ΕΜΠ)

References:

- [1] P. Maragos, V. Charisopoulos, E. Theodosis, "Tropical Geometry and Machine Learning", Proceedings of the IEEE, 109 (5), pp. 728-755, 2021. <https://ieeexplore.ieee.org/document/9394420>
- [2] P. Butkovič. Max-linear systems: theory and algorithms. Springer Science & Business Media, 2010.
- [3] A. Magnani and S. P. Boyd, "Convex piecewise-linear fitting," Optimization and Engineering, vol. 10, no. 1, pp. 1–17, 2009
- [4] P. Maragos and E. Theodosis, "Tropical geometry and piecewise-linear approximation of curves and surfaces on weighted lattices," ArXiv preprint arXiv:1912.03891, 2019
- [5] J. Hook, "Linear regression over the max-plus semiring: algorithms and applications," arXiv preprint arXiv:1712.03499, 2017.
- [6] S. Karaev and P. Miettinen, "Capricorn: An algorithm for subtropical matrix factorization," in Proceedings of the 2016 SIAM international conference on data mining. SIAM, 2016, pp. 702–710
- [7] S. Karaev and P. Miettinen "Cancer: Another algorithm for subtropical matrix factorization," in Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part II 16. Springer, 2016, pp. 576–592
- [8] A. Omanovic, P. Oblak, and T. Curk, "Faststmf: Efficient tropical matrix factorization algorithm for sparse data," arXiv preprint arXiv:2205.06619, 2022.

(*) L7. Μοντέλα Διάχυσης για Χρονοσειρές και Εφαρμογές

Τα τελευταία χρόνια υπάρχει μεγάλο ενδιαφέρον για generative μοντέλα. Μία από τις πιο επιτυχημένες κατηγορίες είναι τα μοντέλα διάχυσης [1-3]. Τα μοντέλα διάχυσης έχουν πρόσφατα επεκταθεί σε δεδομένα χρονοσειρών [4,5]. Αυτό επιτρέπει να δειγματοληπτούμε πιθανές μελλοντικές τροχιές από χρονοσειρές. Μερικά παραδείγματα χρονοσειρών είναι η ζήτηση ενός προϊόντος (π.χ. ενέργειας), η χρήση δεδομένων, η κίνηση στους δρόμους, οι μετρήσεις διαφόρων αισθητήρων κ.λπ.

Στη διπλωματική αυτή προτείνουμε τη μελέτη generative μοντέλων για χρονοσειρές με έμφαση στη χρήση τους για βελτιστοποίηση. Ένα συγκεκριμένο παράδειγμα είναι η αποθήκευση ηλεκτρικής ενέργειας. Προτείνουμε την κατασκευή ενός μοντέλου που να δημιουργεί δείγματα από πιθανές μελλοντικές τροχιές τιμών και προσαρμόζουμε τις τεχνικές για την κατασκευή scenario trees (για τη χρήση τους στη βελτιστοποίηση π.χ. [6]). Αυτό μας επιτρέπει να βελτιστοποιήσουμε τη διαχείριση της αποθήκευσης της ενέργειας. Σαν σύγκριση μπορούμε να έχουμε το όφελος που θα προέκυπτε αν είχαμε γνώση της μελλοντικής τιμής.

(Συνεργάτες: Δρ. Ιωάννης Κορδώνης, ΕΜΠ)

References:

- [1] Sohl-Dickstein, Jascha, et al. "Deep unsupervised learning using nonequilibrium thermodynamics." International conference on machine learning. PMLR, 2015.
- [2] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems 33 (2020): 6840-6851.
- [3] Song, Yang, et al. "Score-based generative modeling through stochastic differential equations." arXiv preprint arXiv:2011.13456 (2020).
- [4] Rasul, K., Seward, C., Schuster, I., & Vollgraf, R. (2021, July). Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In International Conference on Machine Learning (pp. 8857-8868). PMLR.
- [5] Lin, Lequan, et al. "Diffusion models for time series applications: A survey." arXiv preprint arXiv:2305.00624 (2023).
- [6] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. Society for Industrial and Applied Mathematics, 2021.
- [7] Zhang, Xinjing, et al. "Arbitrage analysis for different energy storage technologies and strategies." Energy Reports 7 (2021): 8198-8206.
- [8] Kordonis, Ioannis, Alexandros C. Charalampidis, and Pierre Haessig. "Optimal operation of a grid-connected battery energy storage system over its lifetime." Optimal Control Applications and Methods 44.2 (2023): 739-757.

Επεξεργασία κειμένων και φυσικής γλώσσας:

(* T1. Learning from Structure: Annotation-free Text Recognition.

Χρήση της κατανομής λέξεων κειμένου για εκπαίδευση συστήματος αναγνώρισης χειρόγραφων λέξεων κειμένου χωρίς δεδομένα επισημείωσης. Συγκεκριμένα έχουμε 2 χώρους: τον χώρο οπτικών περιγραφητών από μεμονωμένες εικόνες χειρόγραφων λέξεων και τον χώρο των λέξεων σε μορφή γραμματοσειρών, όπου λέξεις με μικρές διαφορές, π.χ. έναν χαρακτήρα διαφορά: "true" και "tree", πρέπει να βρίσκονται κοντά. Επιθυμούμε να ταιριάξουμε αυτούς τους δύο χώρους βασιζόμενοι στην χρήση της συχνότητας εμφάνισης συγκεκριμένων λέξεων (λέξεις όπως "and", "the" είναι συχνές σε οποιοδήποτε κείμενο) αλλά και της εσωτερικής δομής τους (π.χ. κοινοί χαρακτήρες) για να περιγράψει μονοσήμαντα η κατανομή των λέξεων. Στην συνέχεια θα γίνει προσπάθεια αντιστοίχισης γραμματοσειρών λέξεων με περιγραφητές εικόνων λέξεων. Συνολικά το συγκεκριμένο πρόβλημα είναι μια εφαρμογή domain adaptation (π.χ. optimal transport).

(Συνεργάτες: Δρ. Γιώργος Ρετσινάς, ΕΜΠ)

References:

- [1] G. Retsinas, G. Sfikas and P. Maragos, ["From Seq2Seq recognition to handwritten word embeddings"](#), BMVC, 2020.
- [2] Xu, Hongteng, et al. "Gromov-wasserstein learning for graph matching and node embedding." International conference on machine learning. PMLR, 2019.
- [3] Ding, Ning, et al. "Source-free domain adaptation via distribution estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

T2. Επεξεργασία κειμένων και φυσικής γλώσσας με αλγεβρικές και γεωμετρικές μεθόδους (vector spaces, graphs).

References:

- [1] D. Widdows, "Geometry and Meaning", Vol. 773. Stanford: CSLI publications, 2004.
- [2] P. Gärdenfors, "Conceptual spaces: The Geometry of Thought", MIT Press, 2004.
- [3] T. Mikolov et al. ["Efficient estimation of word representations in vector space"](#), ICLR 2013.