

Magyarázható képszegmentálás wavelet-hálóval

Lieb Hanna-Georgina, Kaszta Tamás, Csató Lehel

Babeş-Bolyai Tudományegetem, Kolozsvár

hanna.lieb@ubbcluj.ro, tamas.kaszta@ubbcluj.ro, lehel.csato@ubbcluj.ro

A mesterséges intelligenciára alapuló modellek elterjedésével egyre fontosabbá válik a modellek *magyarázhatósága*: azaz annak megértése, hogy mi alapján hozzák meg a döntéseiket. Egy ellenőrzőpont bevezetése – esetünkben prototípusok révén – jelentősen segítheti a modellek belső működésének megértését. Jelen munkánkban egy új, magyarázható képszegmentáló modellt mutatunk be, amely waveleteket használ fel a képek jellemzőinek kinyerésére, és ezen jellemzők prototípusokhoz mért távolsága alapján végzi a pixelenkénti osztályozást (szegmentációt). A waveletek révén egy invertálható jellemzőtérbe jutunk, amely lehetővé teszi a prototípusok vizualizációját és értelmezését. A modell teljesítményét a *Cityscapes* adathalmazon értékeltük, amelyen bár nem értünk el kimagasló pontosságot, a modell magyarázhatósága és a prototípusok értelmezhetősége ígéretes eredményeket mutat.

Hivatkozások

- [1] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3213-3223).
- [2] Lieb, H. G., Kaszta, T. (2025) Explainable image segmentation with wavelet-network. Formal Methods and Foundations of Artificial Intelligence, 148.
- [3] Mallat, S. (1999). A wavelet tour of signal processing. Academic Press.
- [4] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247-278.