Prototípus-alapú képszegmentációs modellek értelmezhetőségének vizsgálata

Portik Ábel, Szenkovits Annamária, Bodó Zalán

Babeș–Bolyai Tudományegyetem, Matematika-Informatika kar, Kolozsvár

abel.portik@ubbcluj.ro, annamaria.szenkovits@ubbcluj.ro, zalan.bodo@ubbcluj.ro

A számítógépes látásban alkalmazott mélytanulási modellek megbízható használatához fontos, hogy döntéseik az ember számára is értelmezhetők legyenek [4]. Ebben segíthetnek a prototípus-alapú modellek, amelyek a képek részleteit tanult, jellegzetes mintázatokkal hasonlítják össze.

Kutatásunkban két ilyen képszegmentációs modellel, a ProtoSeg [5] és a ScaleProtoSeg [3] modellekkel foglalkozunk. Az értelmezhetőségük mérésére a szakirodalomban javasolt [1] két mérőszámot, a konzisztenciát és a stabilitást vizsgáljuk. A konzisztencia azt mutatja meg, hogy egy prototípus különböző képeken mennyire következetesen kapcsolódik ugyanahhoz az objektumrészhez, a stabilitás pedig azt, hogy ez a kapcsolat mennyire marad meg, ha zajt adunk a képhez.

Kísérleteinkben két különböző adathalmazon [2] tanulmányozzuk e mérőszámok működését, valamint azt, hogyan befolyásolják őket a választott paraméterek és az adathalmazok minősége. Célunk megérteni, mennyire adnak megbízható képet ezek a mérőszámok a modellek értelmezhetőségéről, feltárni a robusztusságukkal kapcsolatos problémákat, és javaslatokat tenni ezen problémák megoldására.

Hivatkozások

[1] Qihan Huang, Mengqi Xue, Wenqi Huang, Haofei Zhang, Jie Song, Yongcheng Jing, and Mingli Song. Evaluation and improvement of interpretability for self-explainable part-prototype networks, 2023.

[2] Panagiotis Meletis, Xiaoxiao Wen, Chenyang Lu, Daan de Geus, and Gijs Dubbelman. Cityscapes-panoptic-parts and pascal-panoptic-parts datasets for scene understanding, 2020.

[3] Hugo Porta, Emanuele Dalsasso, Diego Marcos, and Devis Tuia. Multi-scale grouped prototypes for interpretable semantic segmentation. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2869–2880. IEEE, February 2025.

[4] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, 2019.

[5] Mikołaj Sacha, Dawid Rymarczyk, Łukasz Struski, Jacek Tabor, and Bartosz Zieliński. Protoseg: Interpretable semantic segmentation with prototypical parts. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1481–1490, 2023.