Abstract

Human visual system is modeled in engineering field providing feature-engineered methods which detect contrasted, surprising, or unusual data into images. This data is "interesting" for humans and leads to numerous applications. Deep learning (DNNs) drastically improved the algorithms efficiency on the main benchmark datasets. However, DNN-based models are counter-intuitive: surprising or unusual data is by definition difficult to learn because of its low occurrence probability. In reality, DNNs models mainly learn top-down features such as faces, text, people, or animals which usually attract human attention, but they have low efficiency in extracting surprising or unusual data in the images.

In this thesis, we propose a model called DeepRare (DR) family model including DeepRare2019 (DR19) and DeepRare2021 (DR21) which uses the power of DNNs feature extraction and the genericity of feature-engineered algorithms. This algorithm is an evolution of a previous version, DR19. DR21 1) does not need any training and uses the default ImageNet training, 2) it is fast even on CPU, 3) our tests on four very different eye-tracking datasets show that DR21 is generic and is always in the within the top models on all datasets and metrics while no other model exhibits such a regularity and genericity. Finally DR21 provides 4) explanation and transparency on why parts of the image are the most surprising at different levels despite the us of a DNN-based feature extractor and 5) it is tested with several network architectures such as VGG16 (V16), VGG19 (V19) and MobileNetV2 (MN2). DeepRare2021 code can be found at VisualAttention-RareFamily.

Keywords: Visual attention prediction, Top-down information, Bottom-up information, Object detection, Face detection, Text detection, Saliency, Rarity, Eye tracking, Deep features, Odd one out, Visibility.