FAME: Feature Activation Map Explanation on Image Classification and Face Recognition

—

Supplemental Material

Xinyi Zhang Manuel Günther
Department of Informatics, University of Zurich
xinyi.zhang@uzh.ch, siebenkopf@googlemail.com

1. Parameter Sensitivity and Optimization

This section provides additional observations on how several hyperparameters influence the runtime and stability of our visualization procedure. In particular, we examine the effect of larger step sizes and reduced iteration numbers. These experiments show that faster configurations can often be used with minimal impact on the visual quality of the generated explanations.

The visualization results in Fig. 1 show how different iteration numbers and step sizes influence the explanations produced by FAME when no early-stopping criterion is applied. Fig. 1(a) illustrates how the FAME visualization evolves across a wide range of iteration numbers. With only 1 iteration, the explanation is extremely coarse and often lacks a meaningful localization on the object. As the number of iterations increases, the heatmaps quickly become sharper and more focused, stabilizing between roughly 75 and 200 iterations. Beyond 300 iterations, the changes are minor, and additional optimization mainly introduces more background activation rather than improving object localization. This analysis highlights an important limitation of FGGB: the original FGGB method uses only a single step of backward propagation, which is insufficient to obtain a detailed or reliable attribution map, as our results clearly show. In contrast, FAME benefits from multiple iterations, allowing it to refine the explanation and produce significantly more precise and interpretable visualizations.

Similarly, varying the step size in Fig. 1(b) affects the strength and smoothness of the activation: smaller η yield smoother but weaker maps, while moderately larger step sizes produce stronger and more contrasted explanations. Very large step sizes may introduce slight instability or noisy patterns in the background, although the main highlighted regions remain consistent.

Table 1: RUNTIME EVALUATION. In (a), we provide the runtime of the FAME method for different numbers of optimization iterations using a ResNet50 model on 1000 ImageNet images. In (b), we compare the runtime of different XAI methods for face recognition attribution on the 700 image pairs contained in the CFP-FP protocol using the IResNet101 model. For fair comparison, all methods are run in a sequential manner.

(a) FAME Iterations

Iterations	Runtime (s)
1	308.35
25	733.95
50	1231.98
75	1678.72
100	2043.20
200	3975.97
300	5891.13
400	7626.18
500	9552.18

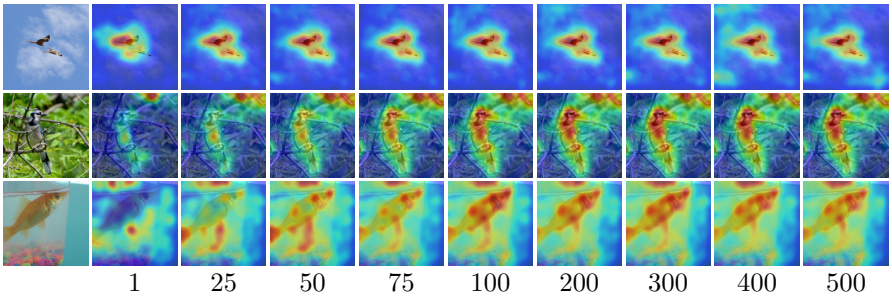
(b) Comparison

XAI	Runtime (s)
Grad-CAM	366.15
Grad-CAM-EW	455.55
CorrRISE	7426.25
FGGB	9558.98
FAME	2116.44

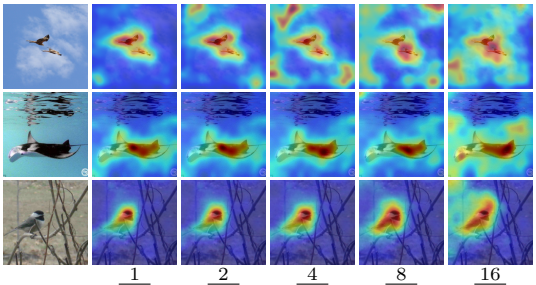
2. Runtime Evaluation

Tab. 1(a) provides timing information obtained when visualizing 1000 images with a ResNet50 backbone. The computation time increases approximately linearly with the number of iterations. Even though larger iteration numbers lead to longer runtimes, our qualitative analysis indicates that meaningful visualizations can already be obtained with considerably fewer iterations, offering a practical trade-off between attribution quality and computational cost. Loss-dependent early-stopping criteria might be applied to stop the iteration without losing visualization power.

In Tab. 1(b), we further report the runtime comparison of all evaluated XAI techniques on face recognition attribution, using the 700 pairs contained in the CFP-FP protocol. Grad-CAM and Grad-CAM-EW remain the fastest methods due to their single-pass backward computation that



(a) Iterations



(b) Step Size η

Figure 1: VISUALIZATIONS WITH FAME PARAMETERS. The figure shows visualization results of the FAME method, when using (a) different numbers of iterations with a fixed step size $\eta = \frac{1}{255}$, and (b) different step sizes η with a fixed number of 100 iterations.

only goes back to the activation map. CorrRISE and FGGB are significantly slower because they rely on repeated masking or iterative gradient computations for elements in the embeddings. Our proposed FAME method with 500 iterations achieves a favorable balance between computational cost and attribution quality. Although slower than Grad-CAM-based approaches, it is substantially faster than CorrRISE and FGGB while delivering the strongest quantitative performance. This confirms that FAME provides an attractive trade-off, offering high-quality explanations with reasonable execution time for real-world verification analysis.

For a fair comparison with existing method implementations that are not parallelized, all visualizations as reported in Tab. 1 are computed sequentially. Our parallelized batch implementation can further increase speed when several visualizations need to be performed at the same time.

3. Visual Results

3.1. Feature Map Visualizations

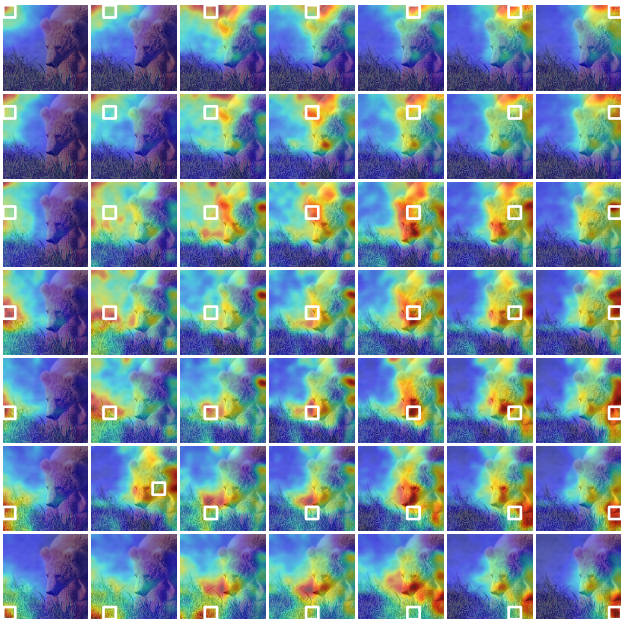
In Figure 1(a) of the main paper, we showed feature maps from one image classification network. In Fig. 2, we show the complete results for all feature map elements for three different networks on ImageNet. Similarly, in Figure 1(b) in the main paper, we showed only one row for visualizing the feature maps of a three face recognition networks. Fig. 3 provides the remaining feature map locations.

3.2. ImageNet Visualizations

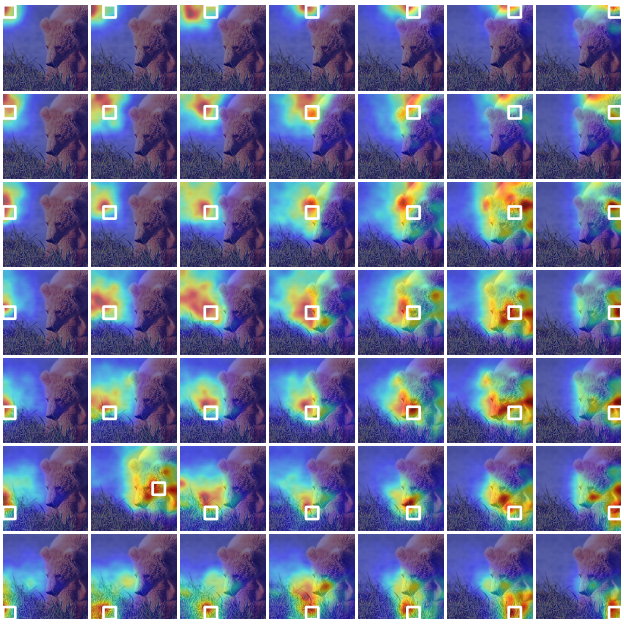
Figure 2 in the main paper showed some visual results for Grad-CAM-EW and FAME for different networks on ImageNet. Fig. 4 shows more examples. As shown in Fig. 4, FAME produces spatially coherent and semantically focused attribution maps across different network architectures, including ResNet34, VGG19, and ConvNeXt-Tiny. Compared with Grad-CAM, FullGradCAM, and HiResCAM, FAME consistently highlights the most discriminative object regions with clearer boundaries and reduced background noise. Importantly, the qualitative patterns remain stable across architectures with substantially different design principles, ranging from classical convolutional networks (VGG19) to residual networks (ResNet34) and modern convolutional-transformer hybrids (ConvNeXt-Tiny). This demonstrates that FAME does not rely on specific feature map structures and generalizes robustly across diverse backbone designs, yielding reliable and architecture-agnostic visual explanations.

3.3. Face Recognition Visualization

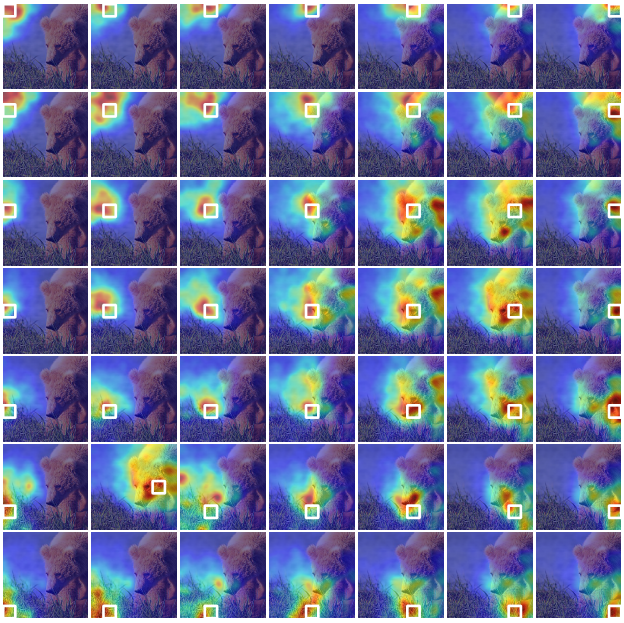
Figure 3 in the main paper included face recognition examples from one network evaluated on three different datasets with one evaluation protocol each. Fig. 5, 6 and 7 include more networks and more protocols.



(a) **ResNet34**

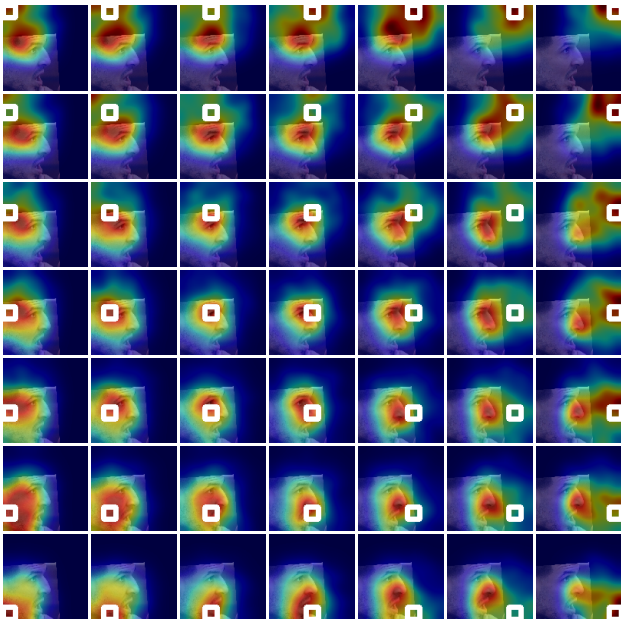


(b) **ResNet50**

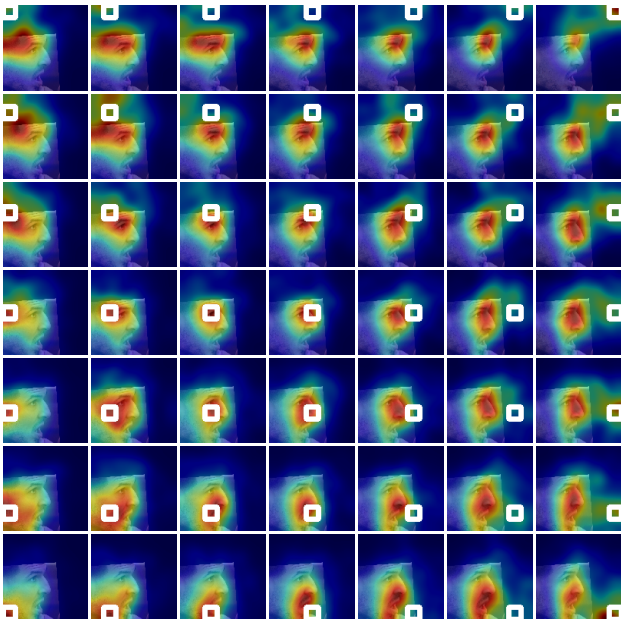


(c) **ResNet101**

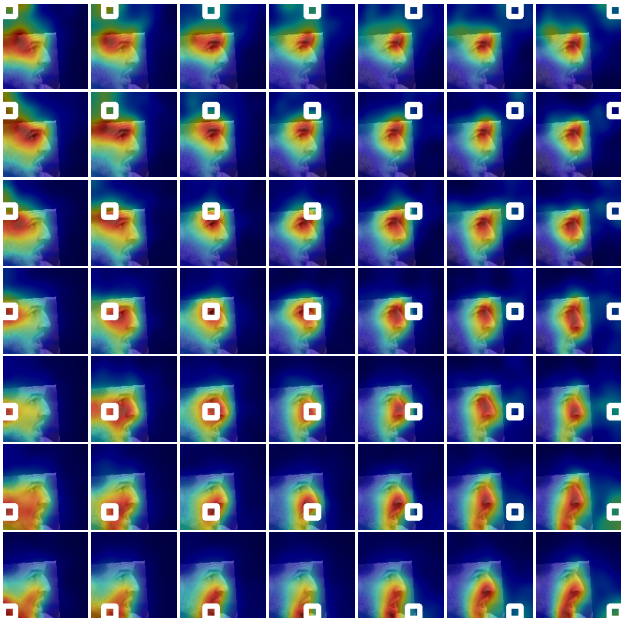
Figure 2: FEATURE MAP VISUALIZATION FOR IMAGENET CLASSIFICATION. The figure shows FAME visualizations of the corresponding regions that are activated for all feature map locations in three image classification networks.



(a) **IResNet18**



(b) **IResNet50**



(c) **IResNet101**

Figure 3: FEATURE MAP VISUALIZATION FOR FACE RECOGNITION. The figure shows FAME visualizations of the corresponding regions that are activated for all feature map locations in three face recognition networks.

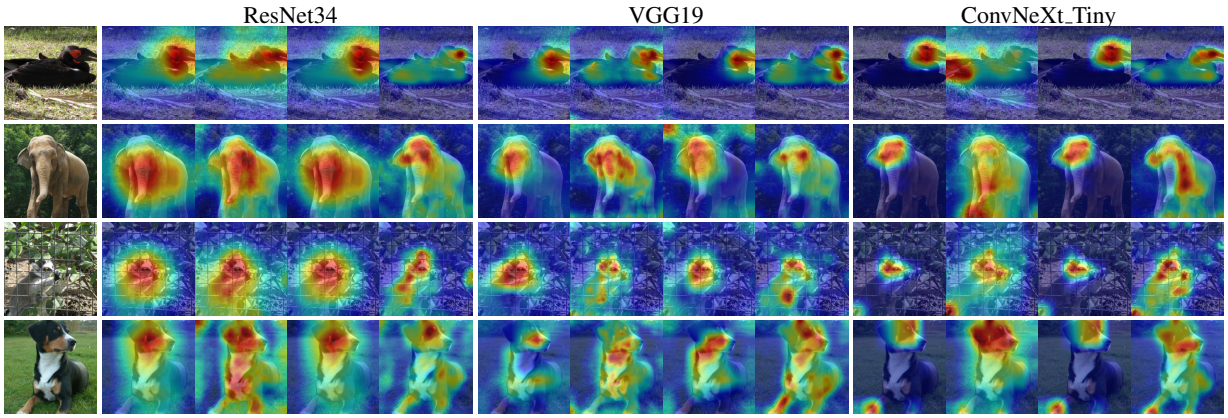
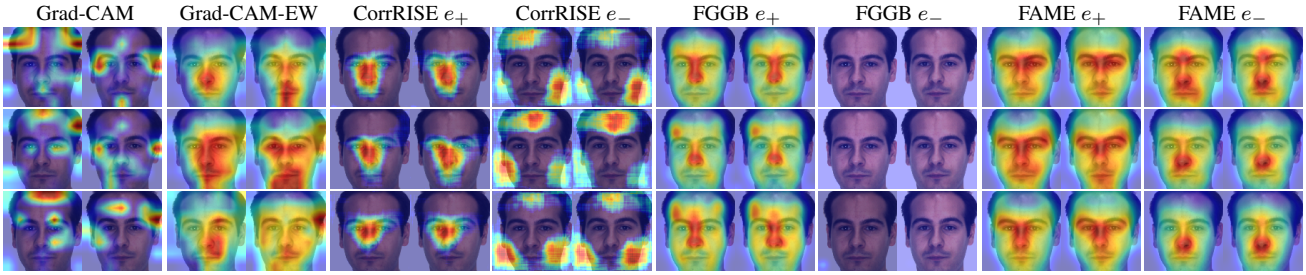
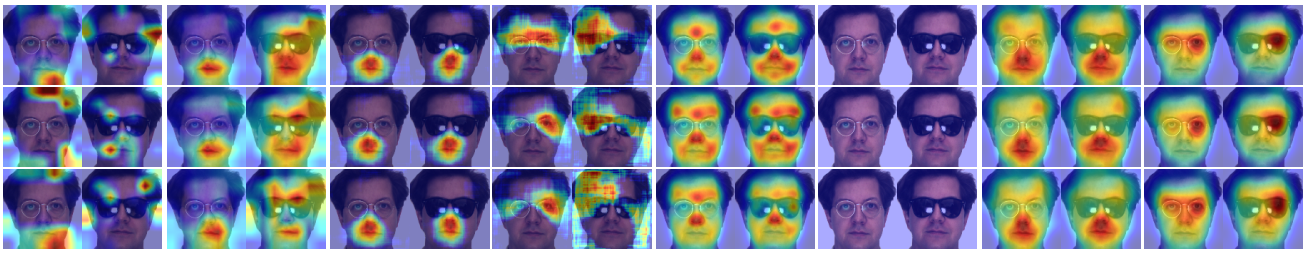


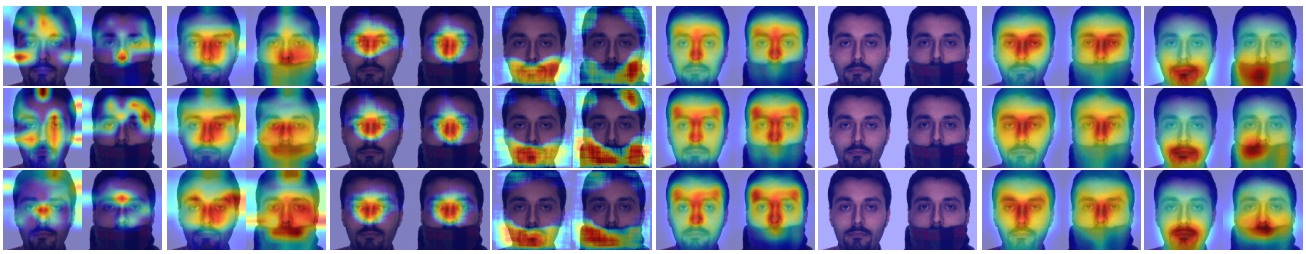
Figure 4: IMAGENET VISUALIZATION. The figure shows the saliency maps generated by (from left to right) Grad-CAM, FullGrad-CAM, HiResCAM and FAME using different models on four ImageNet samples, evaluated with three different pre-trained networks.



(a) Protocol `neutral`

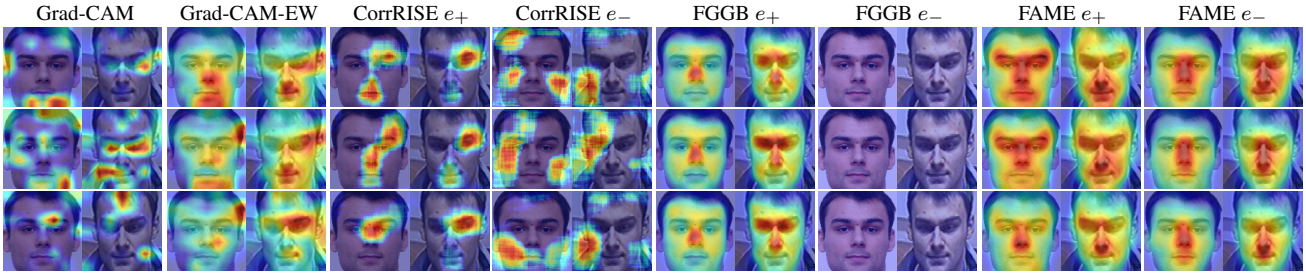


(b) Protocol `glass`

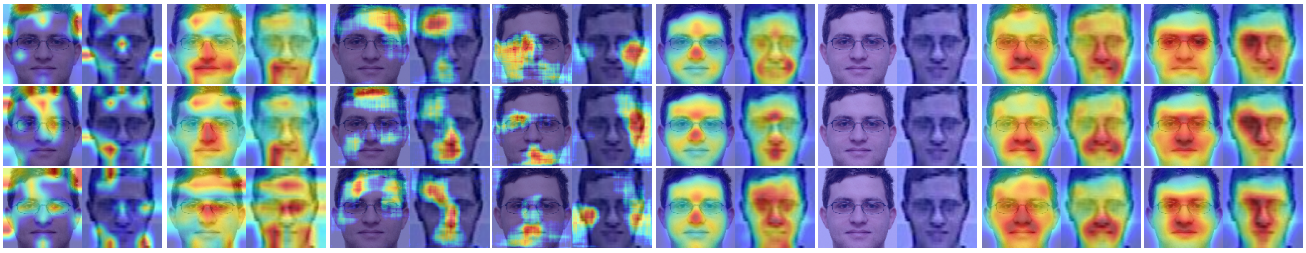


(c) Protocol `scarf`

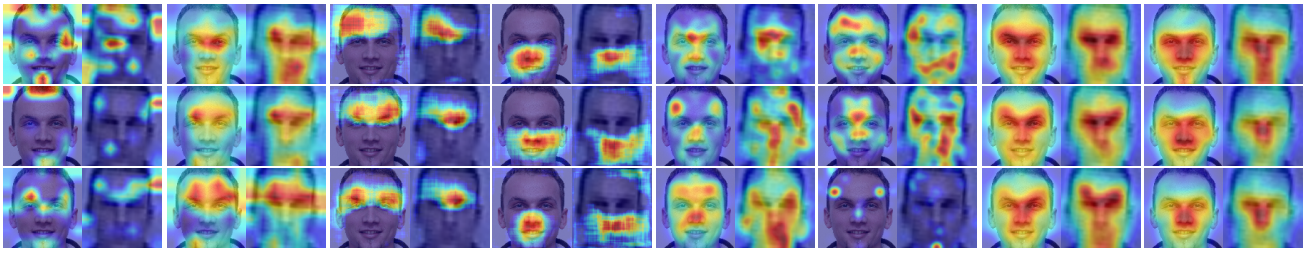
Figure 5: ARFACE. The figure shows a comparative visualization of explanation maps generated on AR face for genuine (same-identity) image pairs using three networks IResNet18, IResNet50 and IResNet101. XAI techniques include Grad-CAM, Grad-CAM-EW, CorrRISE, FGGB, and FAME, including similar e_+ and dissimilar attribution e_- where appropriate.



(a) Protocol `close`

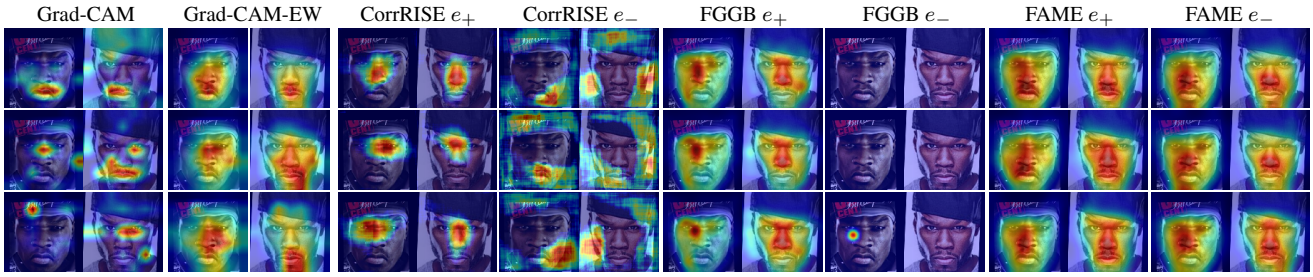


(b) Protocol `medium`

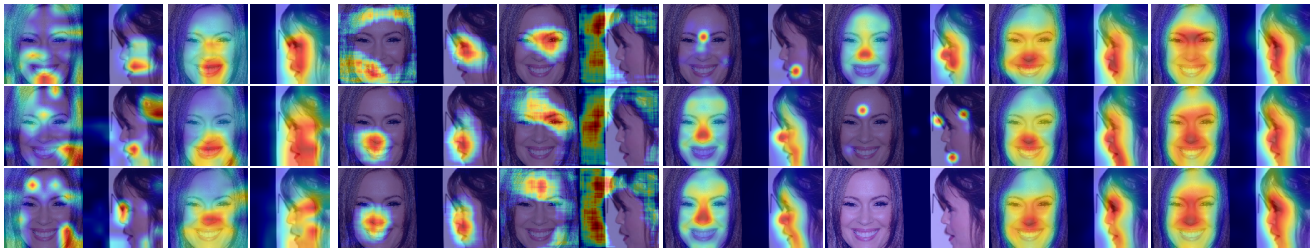


(c) Protocol `far`

Figure 6: SCFACE. The figure shows a comparative visualization of explanation maps generated on SCface for genuine (same-identity) image pairs using three networks IResNet18, IResNet50 and IResNet101. XAI techniques include Grad-CAM, Grad-CAM-EW, CorrRISE, FGGB, and FAME, including similar e_+ and dissimilar attribution e_- where appropriate.



(a) Protocol FF



(b) Protocol FP

Figure 7: CFP. The figure shows a comparative visualization of explanation maps generated on CFP dataset for genuine (same-identity) image pairs using three networks IResNet18, IResNet50 and IResNet101. XAI techniques include Grad-CAM, Grad-CAM-EW, CorrRISE, FGGB, and FAME, including similar e_+ and dissimilar attribution e_- where appropriate.