FedPIN-ICML2024-Rebuttal

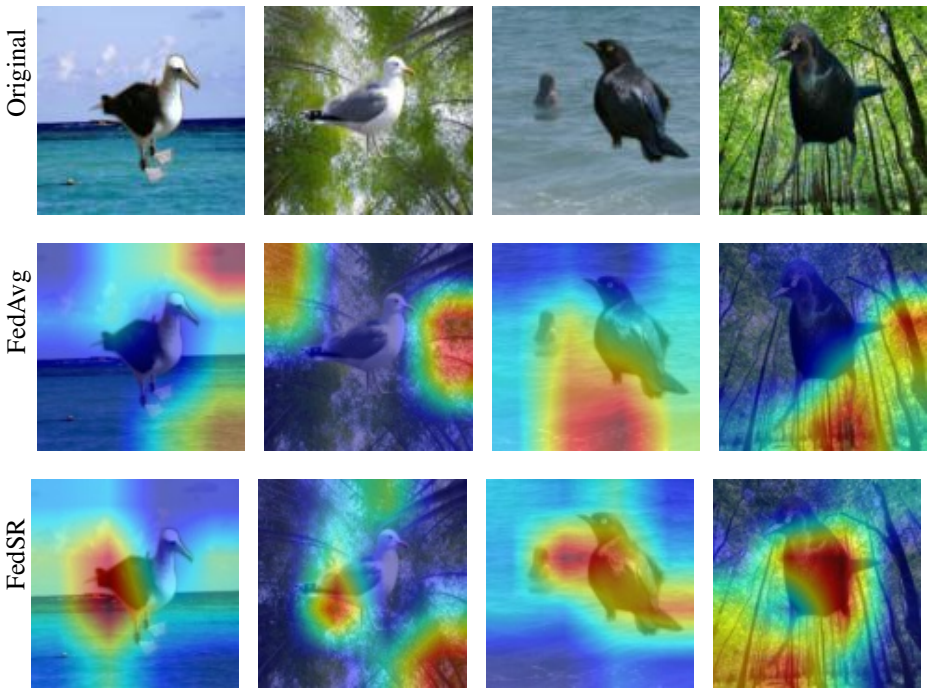
Anonymous Authors[1]

1. I think it would be interesting to look at what are the invariance features and spurious features learned by the model. E.g. for the image data set, would the invariance feature somehow be more related to the shape of the object, not the background?

Answer: To verify that the personalized models developed by our method **FedPIN** rely on the invariant features rather than spurious features, we randomly select one of the obtained personalized models and generate visual explanations for the selected model using Grad-CAM[1]. The commonly used Grad-CAM can produce a localization map which highlights the important regions in the input image for predicting the label. The visualization results of different federated learning (FL) and personalized FL methods for ResNet-18 network on WaterBird dataset are displayed in the following figure. Note that the red regions corresponds to high score for the predicted class. Figure best viewed in color.

The visualization results support that the personalized invariant features extracted by our method FedPIN are more related to the shape of the object, instead of the background.

[1] Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. 2017.



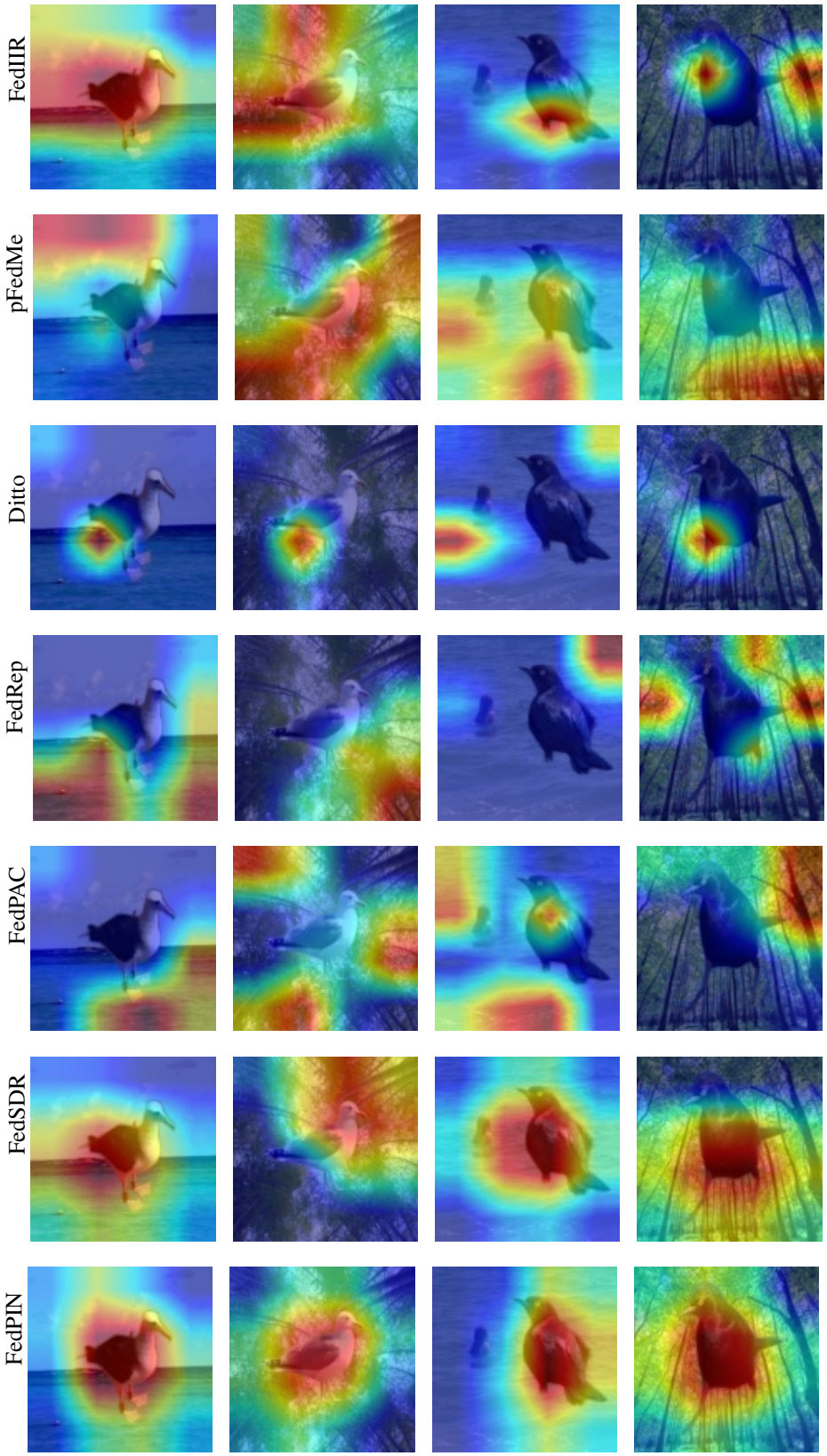


Figure 1. The visualization results of different federated learning (FL) and personalized FL methods on WaterBird dataset are generated by using Grad-CAM[1]. The red regions in the figures correspond to high score for the predicted class. Figure best viewed in color.