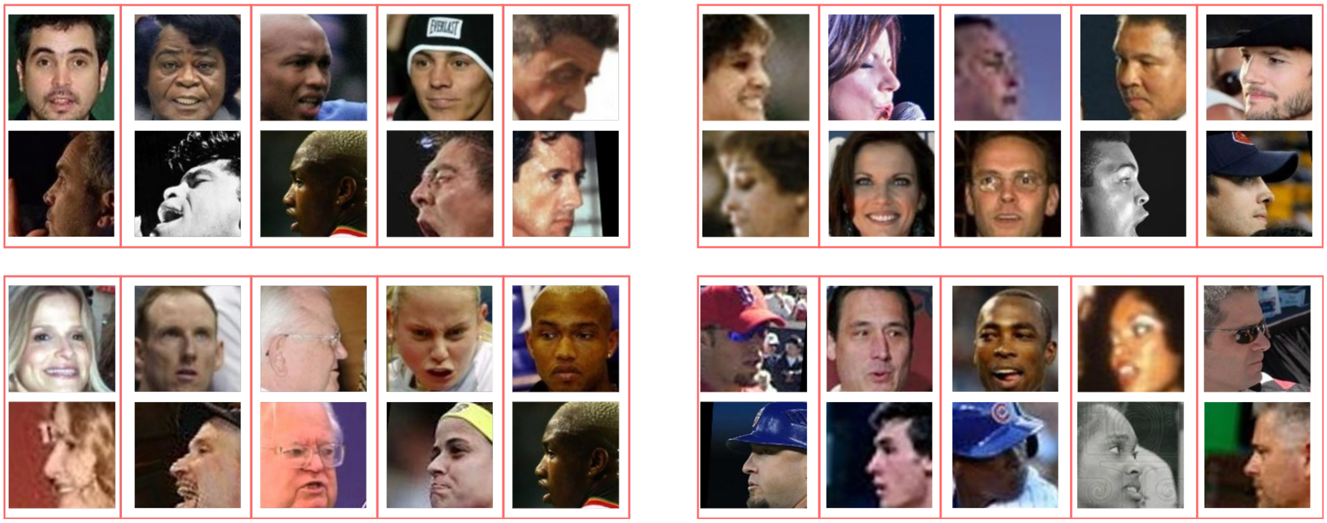
Our Deep CNN Face Matchers Have Developed Achromatopsia
Supplementary Material

Aman Bhatta[1] Domingo Mery[2] Haiyu Wu[1] Joyce Annan[3] Michael C. King[3]
Kevin W. Bowyer[1]

[1]University of Notre Dame
[2]Pontificia Universidad Católica de Chile
[3]Florida Insitute of Technology

A. Is there a reason for why accuracy on CPLFW is low for Grayscale model?

In our experiments, we have found that the difference between RGB (3-channel) and grayscale (1-channel) inputs has the most significant impact on CPLFW dataset for both the AdaFace and ArcFace loss functions. The differences between the RGB and grayscale trained models on the CPLFW dataset for ArcFace and AdaFace are 0.35% and 0.4% respectively. To investigate the reasons for these differences, we utilized a trained ArcFace model and visualized the image pairs in two segments: i) where the model trained on color correctly identified the images while the gray model made errors, and ii) where the model trained on gray correctly identified the images while the color model made errors. This analysis will help to identify any consistent patterns causing discrepancies between the models trained on color and grayscale inputs. In total, there were 346 error pairs for the RGB model and 371 error pairs for the Grayscale model. However, strikingly, 317 error pairs were common to both the color and grayscale models. The exclusive error pairs are displayed in Figure 1 below.



(a) **Grayscale model ✓ Color model ✗**

(b) **Color model ✓ Grayscale model ✗**

Figure 1. **Error pairs exclusive to RGB and Grayscale models on CPLFW dataset.** Each rectangular box displays error pairs, with the top and bottom representing individual images.

Analysis: Figure 1a shows the error pairs where the grayscale model got it right, but color model got it wrong. Figure 1b shows the error pairs where the color model got it right, but grayscale model got it wrong. Based on the error pairs presented in Figure 1a and 1b, **there are no consistent or obvious patterns that caused the models trained on color or grayscale to be incorrect.**

B. RGB Values Representation for Visible Skin Pixels for Identities of Different Races in WebFace4M

B.1. Image Selection

To select suitable quality images for analysis of the face skin region in the image, we use a BiSeNet semantic segmentation [2] to filter out faces where less than 30% of the face area is classified as skin. This gives us mostly frontal faces, without too much occlusion by glasses or scalp hair. To avoid inconsistencies stemming from mouth open/closed and facial expression, we focus on the part of the face above the upper lip, and on the pixels classified as skin (omitting eyebrows and eyes), and calculate the average RGB of skin pixels. For each of the four identities, we select the 50 images that have the largest number of such skin pixels and plot the average RGB for each of their images in a 3D RGB space. If color is useful to separate images belonging to different identities in the training set, then the 50 images of each identity should form a compact cluster that is well separated from the other identities.

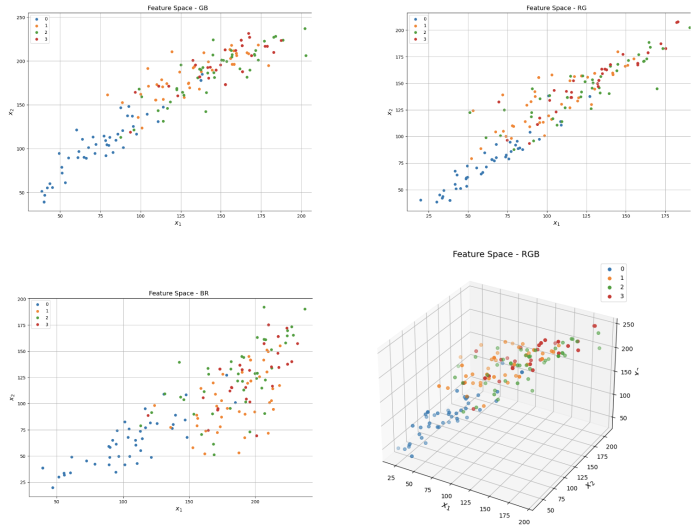


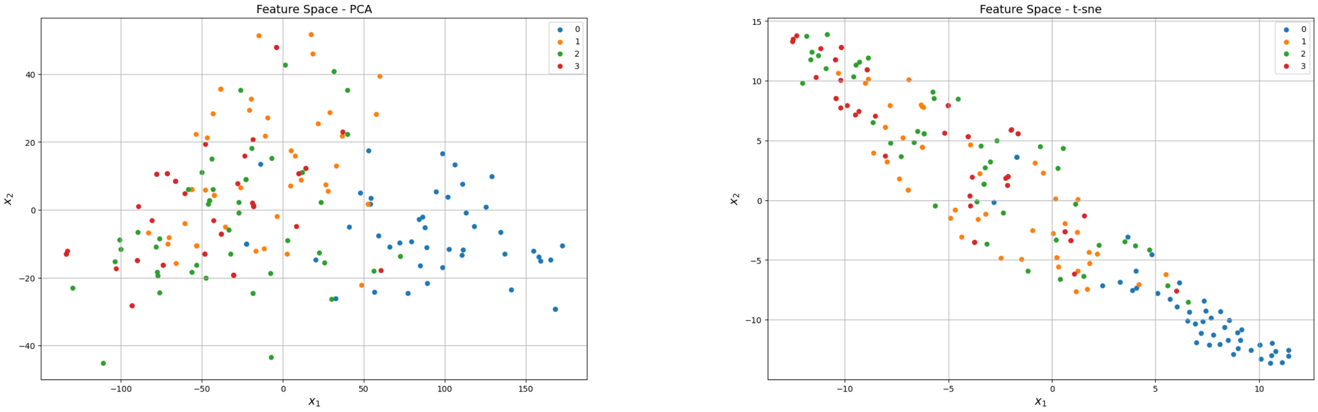
Figure 2. **Multiple Viewpoints of RGB values of visible skin pixels of multiple images for same identity of different ethnicity in WebFace4M**. The figure in the upper-left corner demonstrates the visualization of identities using Green-Blue (GB) coordinates for individuals of diverse ethnicities (the images and identities used are same as in the main paper). The top-right figure exhibits the visualization of the same images using Red-Green (RG) coordinates. Moving to the lower-left side, you'll find the visualization using Blue-Red (BR) coordinates for those very images. Lastly, the figure on the bottom-right showcases the 3D plane visualization. *None of the view shows a tight clustering for visible skin value pixels for same identity*.

B.2. Visualizations

In the main paper, we displayed an "optimal" viewpoint that presents the greatest separation within the RGB space. As depicted in the main paper's figure, it is clear that the RGB representation of visible skin pixels for a single identity is not closely clustered. To provide additional proof of this lack of tight clustering in the RGB representation for the same identity, we exhibit multiple 2D viewpoints (BG, RB, RG) and an alternative 3D representation for skin pixels in Figure 2.

C. t-SNE and PCA of RGB values of visible skin pixels for Identities of Different Races in WebFace4M

In the previous section, we exhibited a straightforward color-coordinate representation for the RGB values of visible skin pixels. Within this section, we introduce the t-SNE and PCA representations of these RGB values for visible skin pixels. Figure 3a and 3b reveal that there isn't a distinct clustering of points for numerous images of the same identity.



(a) PCA

(b) t-SNE

Figure 3. **t-SNE and PCA representation of RGB co-ordinates of visible skin pixels for identities in WebFace4M**

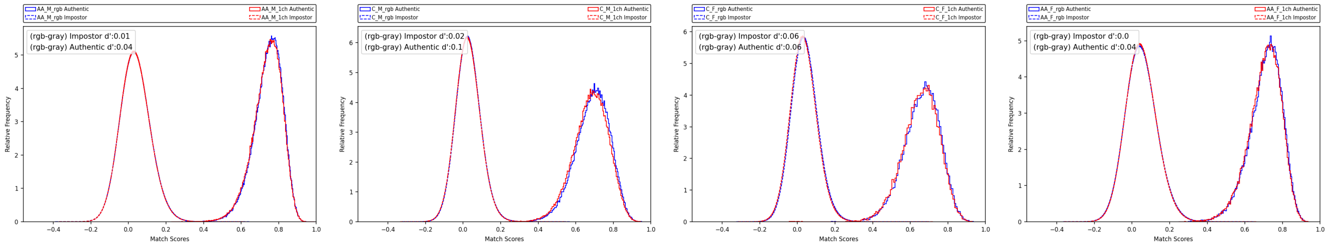
D. Accuracy of model trained with one-channel grayscale vs three-channel RGB on MORPH

In this section, we present the distribution of genuine and impostor scores for the combined-margin model utilizing the ArcFace loss. We examine two training instances: a) A model trained with one-channel grayscale images, and b) A model trained with three-channel RGB images. The impostor set comprises similarity scores obtained from image pairs representing different identities, while the genuine set consists of similarity scores from image pairs representing the same identity. To assess the separation between two distributions, we employ the d-prime (d') metric [1]. The d-prime metric is calculated as follows:

$$d' = \frac{|\mu_{D1} - \mu_{D2}|}{\sqrt{\frac{\sigma_{D1}^2 + \sigma_{D2}^2}{2}}} \tag{1}$$

where μ_{D1} and σ_{D1} are the mean and standard deviation for the distribution 1, and μ_{D2} and σ_{D2} for distribution 2. Larger d' means greater separation between two distributions, generally implying better accuracy if one of the distribution represents impostor score and other represents genuine scores. (Using d' assumes that the distributions are reasonably approximated as Gaussian.)

Figure 4 illustrates the genuine and impostor distributions for different demographic groups, including Caucasian Male (C M), Caucasian Female (C F), African-American Male (AA M), and African-American Female (AA F). The d' prime values, presented as unsigned values, indicate the comparison of distances between three-channel RGB and one-channel grayscale. Notably, all d' values are either positive or zero for impostor comparisions, suggesting that the impostor distribution is marginally better or equal for one-channel comparison compared to RGB (where lower scores indicate better impostor distributions). However, for the genuine distribution, RGB consistently exhibits a slight advantage (where higher scores indicate better genuine distributions).



(a) AA M

(b) C M

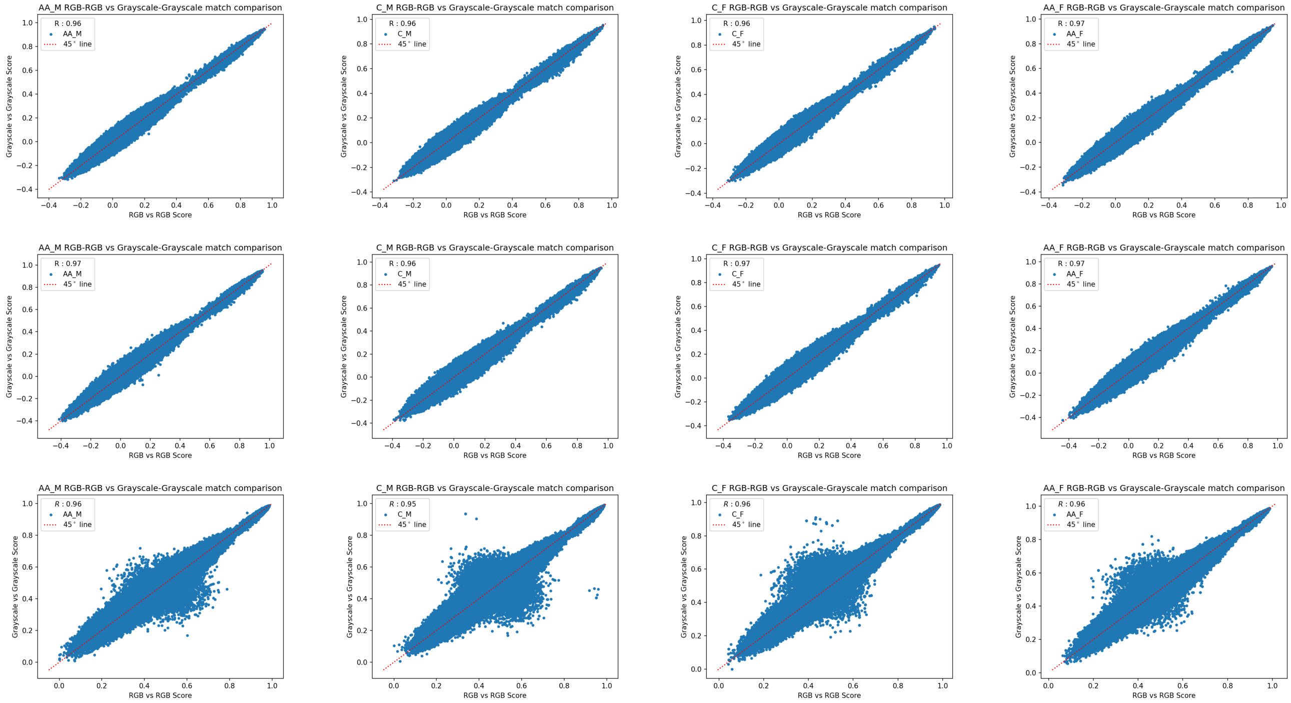
(c) C F

(d) AA F

Figure 4. **Genuine and Impostor Distributions for models trained with three RGB image vs one-channel grayscale images**. The plot shows the comparison of the performance of ArcFace models trained on three-channel RGB images versus one-channel grayscale images. It shows that both models have similar performance when tested on the MORPH dataset.

E. RGB-RGB vs Grayscale-Grayscale scatter plots on pre-trained models

The main paper features a scatter plot that illustrates the cosine similarity between image pairs in RGB format on the horizontal axis and the similarity between their grayscale versions on the vertical axis, for the combined-margin model with the ArcFace loss. In this section, we extend this analysis to include AdaFace, MagFace, and COTS matcher. By doing so, we demonstrate that the observation we made regarding the relationship between RGB and grayscale similarities is not limited to the ArcFace loss alone. The inclusion of these additional models ensures that our findings hold true across multiple approaches, establishing the generality of our observations. As demonstrated in the main paper, the points in the plot cluster around the 45-degree line, accompanied by a high Pearson product-moment correlation coefficient (R). This observation indicates that the computed similarity remains nearly unchanged whether derived from the color or grayscale version of the image pair.



(a) MORPH AA M

(b) MORPH C M

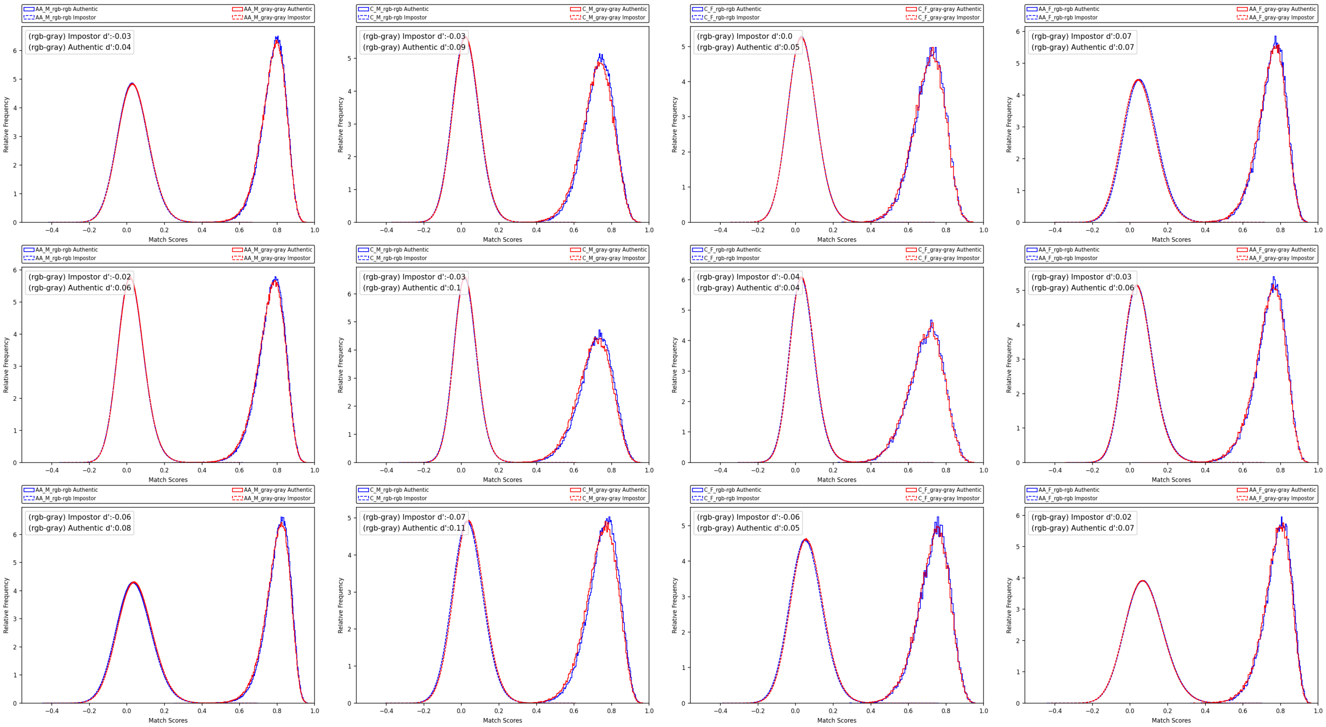
(c) MORPH C F

(d) MORPH AA F

Figure 5. **RGB-RGB vs Grayscale-Grayscale similarity scatter plots for different demographic groups in MORPH**. The top row shows the result for AdaFace, the middle row for the MagFace and the last row for the COTS(Commmercial-off-the-shelf) matcher.

F. RGB-RGB vs Grayscale-Grayscale genuine and impostors on pretrained models

In this section, we present a comparison between RGB-RGB similarity and grayscale-grayscale similarity using several pre-trained models. We examine the genuine and impostor distributions for the MORPH dataset, employing models trained with the ArcFace, AdaFace, and MagFace loss functions. *The MORPH dataset presents a strong test for the possible value of color information, because the images are all mugshot-style images, acquired by an operator under controlled acquisition with consistent lighting and the subject standing in front of a uniform gray background.* The top row of the results showcases the performance of the ArcFace model, followed by the middle row representing the AdaFace model, and finally the bottom row depicting the MagFace model. Each column in the figure represents the impostor-impostor and genuine-genuine comparisons for different demographic groups: African-American male, Caucasian male, Caucasian female, and African-American female, respectively. As highlighted in the main paper and illustrated in Figure 6, it is evident that the three most popular and state-of-the-art CNN-based face embedding networks, namely ArcFace, MagFace, and AdaFace, are unable to leverage color information. Consequently, their performance remains unaffected when comparing RGB and grayscale face images.



(a) MORPH AA M

(b) MORPH C M

(c) MORPH C F

(d) MORPH AA F

Figure 6. **RGB-RGB vs Grayscale-Grayscale genuine and impostor plots for different demographic groups in MORPH**. The top row shows the result for MagFace, the middle row for the AdaFace and the last row for the COTS matcher.

G. Color in First Convolutional Layer Weights

In the main paper, we showcased weight distribution visualizations for two retrained combined margin models based on the ArcFace loss: a) The first model was trained on the “color-cleaned" WebFace4M dataset. b) The second model was trained on the “color-cleaned" WebFace4M dataset using the HSV color space.

In this section, *we extend the analysis by showcasing the weight visualization for the AdaFace model retrained on the “color-cleaned" WebFace4M dataset on RGB color space.* Additionally, we include visualizations for several widely used pre-trained models, namely ArcFace, AdaFace, and MagFace, which are readily available in the model zoo of their respective repositories. These pre-trained models can be easily downloaded and are widely utilized in the face recognition community. Through this analysis, we assure that the weight visualization patterns observed in the retrained models were not specific to our training process but are consistent across the pre-trained models as well. Although the patterns may differ due to the choice of loss function, they remain consistent between the pre-trained and retrained instances. Notably, regardless of the loss function employed, a significant majority of the convolution weights tend to converge to zero values.

G.1. Re-trained Adaface Model with ResNet-50 Backbone trained on “color-cleaned" WebFace4M

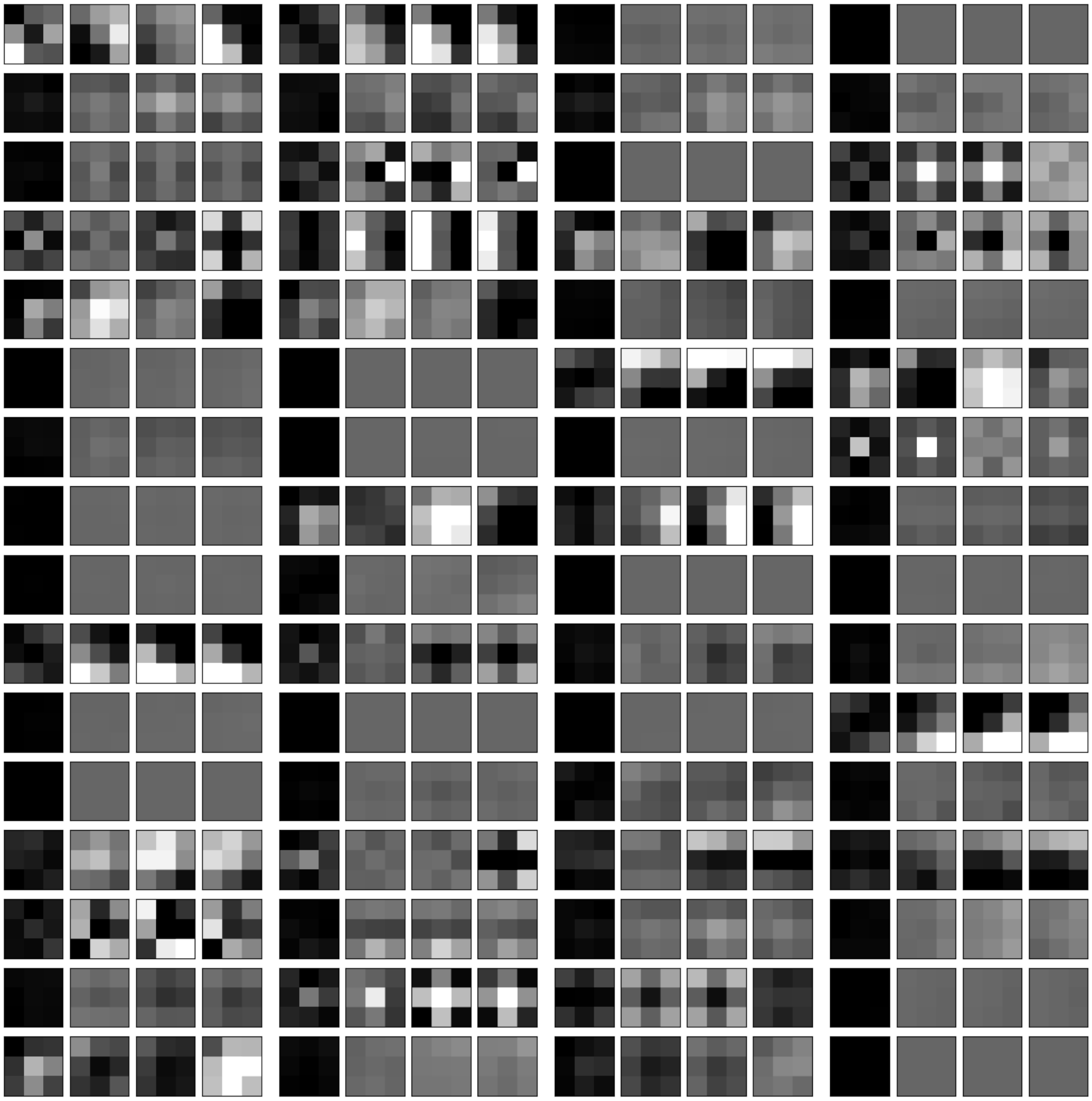


Figure 7. **Visualization of Convolution Filter Weight Values for the First Convolution Block in row-major order.**

G.2. Pre-trained Combined Margin Model (ArcFace loss) with ResNet-100 Backbone trained on Glint-360K

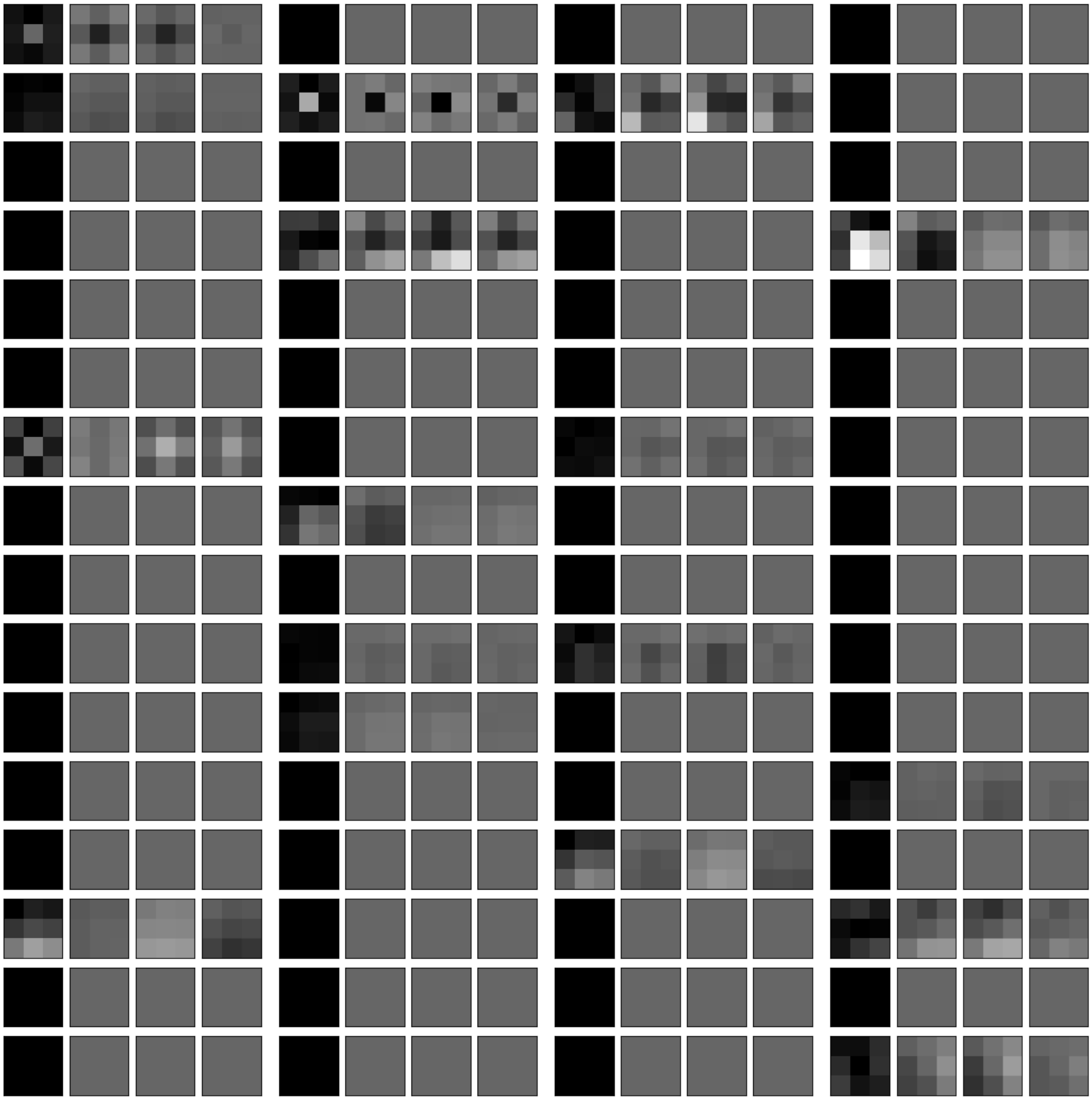


Figure 8. **Visualization of Convolution Filter Weight Values for the First Convolution Block in row-major order.**

G.3. Pre-trained Combined Margin Model (ArcFace loss) with ResNet-100 Backbone trained on MS1MV2

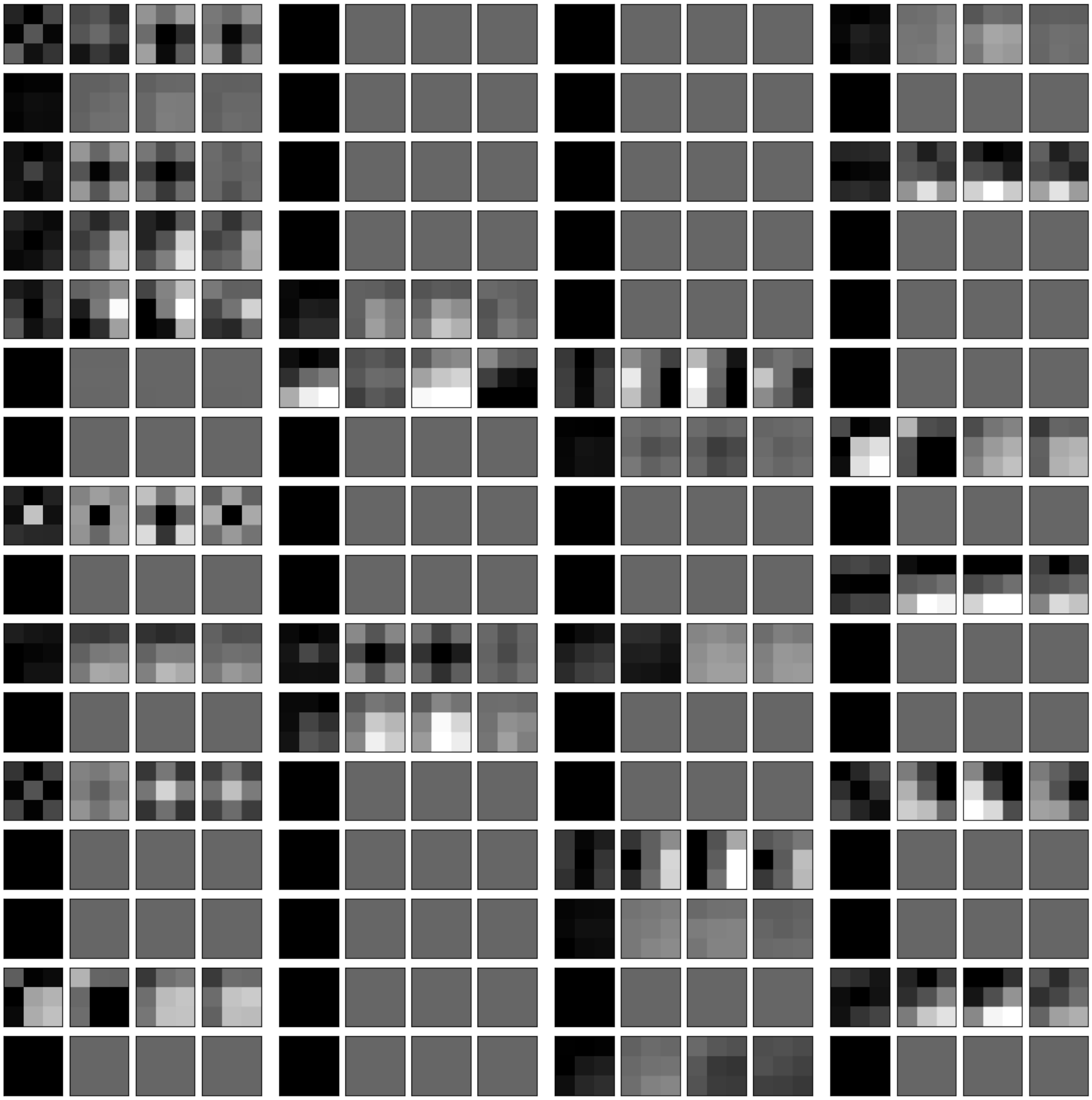


Figure 9. **Visualization of Convolution Filter Weight Values for the First Convolution Block in row-major order.**

G.4. Pre-trained AdaFace with ResNet-100 Backbone trained on MS1MV2

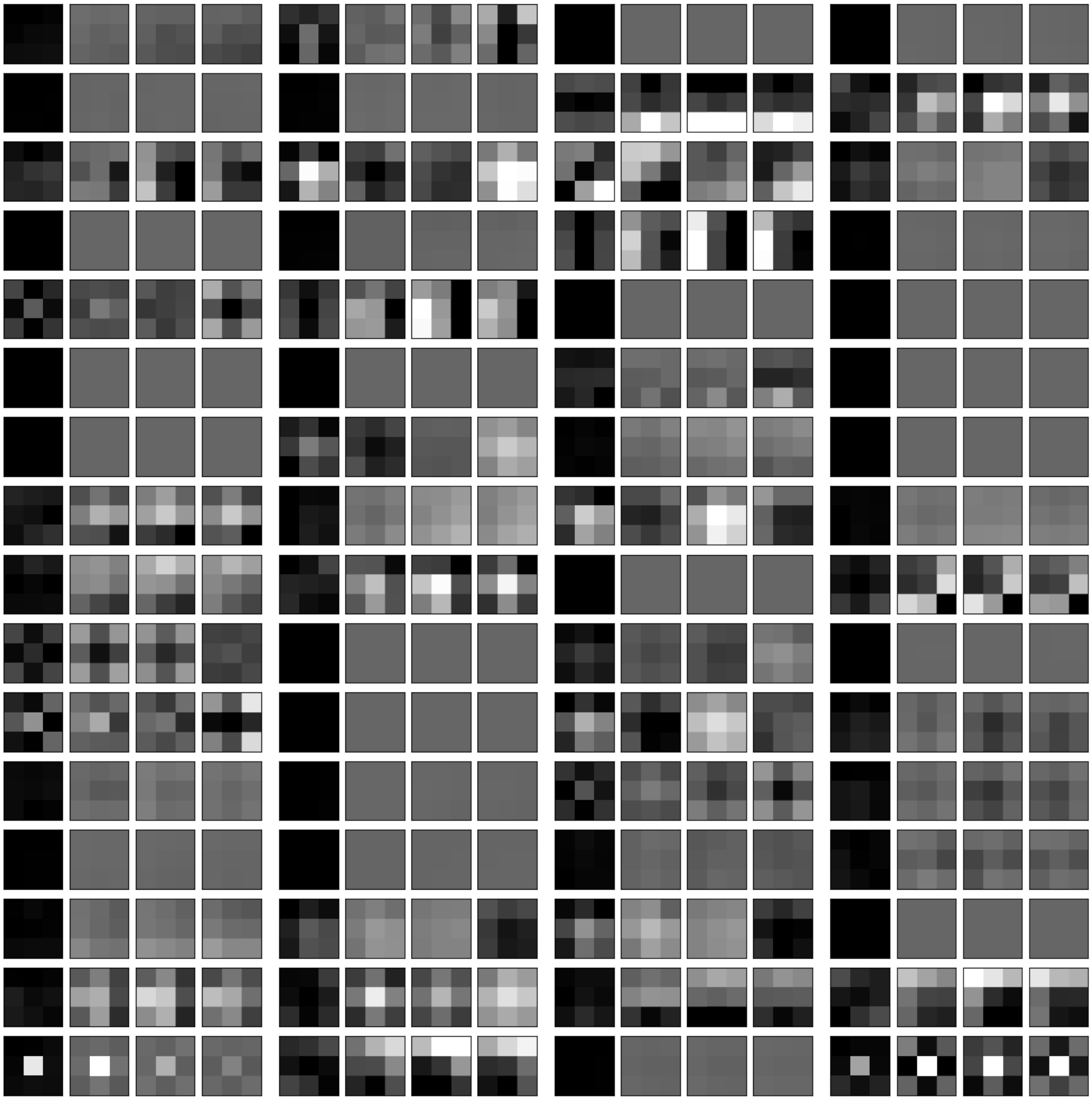


Figure 10. **Visualization of Convolution Filter Weight Values for the First Convolution Block in row-major order.**

G.5. Pre-trained MagFace with ResNet-100 Backbone trained on MS1MV2

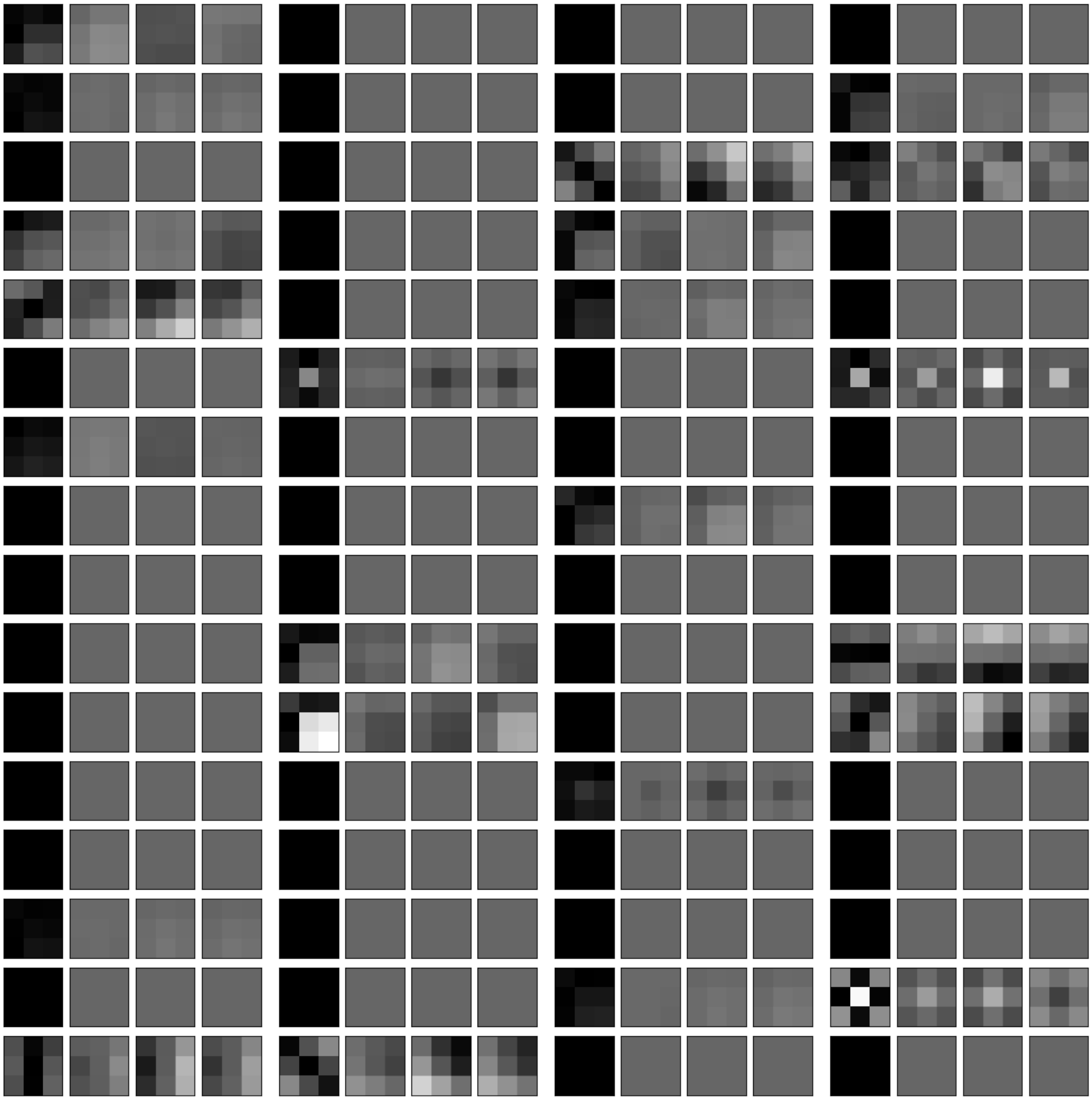


Figure 11. **Visualization of Convolution Filter Weight Values for the First Convolution Block in row-major order.**

References

[1] John Daugman. How iris recognition works. In *The essential guide to image processing*, pages 715–739. Elsevier, 2009. 3

[2] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 325–341, 2018. 2