# Appendix

## FACESEC: A Fine-grained Robustness Evaluation Framework for Face Recognition Systems

Liang Tong[1,2*], Zhengzhang Chen[2†], Jingchao Ni[2], Wei Cheng[2],
Dongjin Song[3], Haifeng Chen[2], Yevgeniy Vorobeychik[1]
[1]Washington University in St. Louis, {liangtong, yvorobeychik}@wustl.edu
[2]NEC Laboratories America, {zchen,jni,weicheng,haifeng}@nec-labs.com
[3]University of Connecticut, dongjin.song@uconn.edu

## A. Grid-level Face Mask Attack

### A.1. Formulation

The optimization formulations of the proposed grid-level face mask attacks under different settings are presented in Table 1. Here, $S$ is the target face recognition model, $\boldsymbol{x}$ is the original input face image. $\boldsymbol{\delta} \in \mathbb{R}^{a\times b}$ is a $a \times b$ color matrix; each element of $\boldsymbol{\delta}$ represents an RGB color. $M$ denotes the mask matrix that constrains the area of perturbation; it contains 1s where perturbation is allowed, and 0s where there is no perturbation. For closed-set systems, $\ell$ denotes the softmax cross-entropy loss function, $y$ is the identity of $\boldsymbol{x}$, and $y_t$ is the target identity for impersonation attacks. For open-set settings, $d$ is the cosine distance (obtained by subtracting cosine similarity from one), $\boldsymbol{x}^*$ is the gallery image of $\boldsymbol{x}$, and $\boldsymbol{x}_t^*$ is the target gallery image for impersonation. $\mathcal{T}$ represents a set of transformations that convert the color matrix $\boldsymbol{\delta}$ to a face mask with a color grid in digital space. Specifically, $\mathcal{T}$ contains two transformations: *interpolation transformation* and *perspective transformation*, which are detailed below.

### A.2. Interpolation Transformation

The interpolation transform starts from a $a \times b$ color matrix $\boldsymbol{\delta}$ and uses the following two steps to scale $\boldsymbol{\delta}$ into a face image, as illustrated in Fig. 1: First, it resizes the color matrix from $a \times b$ to a rectangle $\boldsymbol{\delta}'$ with $c \times d$ pixels, so as to reflect the size of a face mask in a face image in digital space while preserving the layout of the color grids represented by $\boldsymbol{\delta}$. Specifically, in FACESEC, each input face image has $224\times224$ pixels. Let $(a, b) = (8, 16)$ and $(c, d) = (80, 160)$. Then, we put the face mask $\boldsymbol{\delta}'$ into a background image, such that the pixels in the rectangular area have the same value with $\boldsymbol{\delta}'$, and those outside the face mask area have values of 0s.

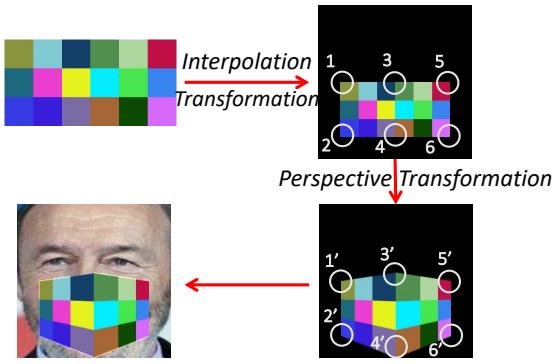


Figure 1. Transformations for the grid-level face mask attack.

### A.3. Perspective Transformation

Once the rectangle $\boldsymbol{\delta}'$ is embedded into a background image, we use a 2-D alignment that relies on the perspective transformation by the following steps. First, we divide $\boldsymbol{\delta}'$ into a left half part $\boldsymbol{\delta}'_L$ and a right half part $\boldsymbol{\delta}'_R$; each is rectangular with four corners. Then, we apply the perspective transformation to project each part to be with aligned coordinates, such that the new coordinates align with the position when a face mask is put on a human face, as shown in Fig. 1. Let $\boldsymbol{\delta}''_L$ and $\boldsymbol{\delta}''_R$ be the left and right part of the aligned face mask, the perspective transformation aims to find a $3 \times 3$ matrix $N_k$ ($k \in \{L, R\}$) for each part such that the coordinates satisfy:

$$\boldsymbol{\delta}''_k(x, y) = \boldsymbol{\delta}'_k(u, v), \quad k \in \{L, R\},$$

where

$$u = \frac{N_k(1,1)x + N_k(1,2)y + N_k(1,3)}{N_k(3,1)x + N_k(3,2)y + N_k(3,3)},$$

and

$$v = \frac{N_k(2,1)x + N_k(2,2)y + N_k(2,3)}{N_k(3,1)x + N_k(3,2)y + N_k(3,3)}.$$

*Work done during an internship at NEC Laboratories America.
†Corresponding author.

Table 1. Optimization formulations of grid-level face mask attacks.

| Target System | Attacker's Goal | Formulation |
|---|---|---|
| Closed-set | Dodging | $\max_{\boldsymbol{\delta}} \ell(S(\boldsymbol{x} + M \cdot \mathcal{T}(\boldsymbol{\delta})), y)$ |
| Closed-set | Impersonation | $\min_{\boldsymbol{\delta}} \ell(S(\boldsymbol{x} + M \cdot \mathcal{T}(\boldsymbol{\delta})), y_t)$ |
| Open-set | Dodging | $\max_{\boldsymbol{\delta}} d(S(\boldsymbol{x} + M \cdot \mathcal{T}(\boldsymbol{\delta})), S(\boldsymbol{x}^*))$ |
| Open-set | Impersonation | $\min_{\boldsymbol{\delta}} d(S(\boldsymbol{x} + M \cdot \mathcal{T}(\boldsymbol{\delta})), S(\boldsymbol{x}_t^*))$ |

**Algorithm 1** Computing adversarial face mask.

**Input:** Target system $S$;
Input face image $\boldsymbol{x}$ and its identity $y$;
The number of iterations $T$;
Step size $\alpha$;
Momentum parameter $\mu$.
**Output:** The color matrix of adversarial face mask $\boldsymbol{\delta}_T$.

Initialize the color matrix $\boldsymbol{\delta}_0 := \mathbf{0}$, momentum $\boldsymbol{g}_0 := \mathbf{0}$;
Use interpolation and perspective transformations to convert $\boldsymbol{\delta}_0$: $\boldsymbol{\delta}_0^{''} := \mathcal{T}(\boldsymbol{\delta}_0)$;
**for** each $t \in [0, T-1]$ **do**
  $\boldsymbol{g}_{t+1} := \mu \cdot \boldsymbol{g}_t + \frac{\nabla_{\boldsymbol{\delta}_t} \ell(S(\boldsymbol{x}+M\cdot\boldsymbol{\delta}_t^{''}), y)}{||\ell(S(\boldsymbol{x}+M\cdot\boldsymbol{\delta}_t^{''}), y)||_1}$;
  $\boldsymbol{\delta}_{t+1} := \boldsymbol{\delta}_t + \alpha \cdot \text{sign}(\boldsymbol{g}_{t+1})$;
  $\boldsymbol{\delta}_{t+1}^{''} := \mathcal{T}(\boldsymbol{\delta}_{t+1})$;
  Clip $\boldsymbol{\delta}_{t+1}^{''}$ such that pixel values of $\boldsymbol{x} + M \cdot \boldsymbol{\delta}_{t+1}^{''}$ are in $[0, 255/255]$;
**end for**
**return** $\boldsymbol{\delta}_T$.

**Algorithm 2** Finding universal perturbations.

**Input:** Target system $S$;
Input face image batch $\{\boldsymbol{x}_i, y_i\}_{i=1}^N$;
The number of iterations $T$;
Step size $\alpha$;
Momentum parameter $\mu$.
**Output:** The universal perturbation $\boldsymbol{\delta}_T$ for $\{\boldsymbol{x}_i, y_i\}_{i=1}^N$.

1: Initialize $\boldsymbol{\delta}_0 := \mathbf{0}$, $\boldsymbol{g}_0 := \mathbf{0}$;
2: **for** each $t \in [0, T-1]$ **do**
3:  **for** each $i \in [1, N]$ **do**
4:  $\ell_{i,t} := \ell(S(\boldsymbol{x}_i + M \cdot \boldsymbol{\delta}_t), y_i)$;
5:  **end for**
6:  $\ell_t = \min\{\ell_{i,t}\}_{i=1}^N$;
7:  $\boldsymbol{g}_{t+1} := \mu \cdot \boldsymbol{g}_t + \frac{\nabla_{\boldsymbol{\delta}_t} \ell_t}{||\ell_t||_1}$;
8:  $\boldsymbol{\delta}_{t+1} := \boldsymbol{\delta}_t + \alpha \cdot \text{sign}(\boldsymbol{g}_{t+1})$;
9:  Clip $\boldsymbol{\delta}_{t+1}$ such that pixel values of $\boldsymbol{x} + M \cdot \boldsymbol{\delta}_{t+1}$ are in $[0, 255/255]$;
10: **end for**
11: **return** $\boldsymbol{\delta}_T$.

Finally, we merge $\boldsymbol{\delta}_L^{''}$ and $\boldsymbol{\delta}_R^{''}$ to obtain the aligned grid-level face mask.

### A.4. Computing Adversarial Face Masks

The algorithm for computing the color grid for adversarial face mask attack is outlined in Algorithm 1. Here, we use the dodging attack on closed-set systems as an example. The algorithms for other settings are similar. Note that $\boldsymbol{\delta}_T$ is the resulting color matrix, and the corresponding adversarial example is $\boldsymbol{x} + M \cdot \mathcal{T}(\boldsymbol{\delta}_T)$.

## B. Universal Attack

### B.1. Optimization Formulation

The formulations of universal perturbations are presented in Table 2. In FACESEC, we mainly focus on universal dodging attacks. Effective universal impersonation attack is still an open problem, and we leave it for future work.

### B.2. Computing Universal Perturbations

The algorithm for finding universal perturbations is presented in Algorithm 2. Here, we use the dodging attack on closed-set systems as an example. The algorithms for other settings are similar. Note that in practice, Line 3–6 in Algorithm 2 can be executed in a paralleled manner by using GPUs. Therefore, compared to traditional methods that iterate every data point to find a universal perturbation [3], our approach can achieve a significant speedup.

## C. Robustness of Face Recognition Components

### C.1. Open-set Systems Under Dodging Attacks

To study the robustness of open-set system components under dodging attacks, we employ six different face recognition systems and then evaluate the attack success rates of dodging attacks corresponding to different target and surrogate face recognition models. Specifically, besides the five systems (VGGFace, FaceNet, ArcFace18, ArcFace50, and ArcFace101) presented in Table 2 of the main paper, we build a face recognition model by training FaceNet [4] using the VGGFace2 dataset [1] (henceforth, *FaceNet+*). Here, FaceNet and FaceNet+ are trained using the same neural architecture but different training sets, while the ArcFace variations share the same training data but with different architectures. The results are presented in Fig. 2.

We have the following two observations, which are sim-

Table 2. Optimization formulations of universal dodging attacks.

| Target System | Perturbation Type | Formulation |
|---|---|---|
| Closed-set | Pixel-level | $\max_{\boldsymbol{\delta}} \min\{\ell(S(\boldsymbol{x}_i + M\boldsymbol{\delta}), y_i)\}_{i=1}^{N}, \;\; s.t. \; \|\|\boldsymbol{\delta}\|\|_p \leq \epsilon$ |
| Closed-set | Grid-level | $\max_{\boldsymbol{\delta}} \min\{\ell(S(\boldsymbol{x}_i + M \cdot \mathcal{T}(\boldsymbol{\delta})), y_i)\}_{i=1}^{N}$ |
| Open-set | Pixel-level | $\max_{\boldsymbol{\delta}} \min\{d(S(\boldsymbol{x_i} + M\boldsymbol{\delta}), S(\boldsymbol{x}_i^*))\}_{i=1}^{N}, \;\; s.t. \; \|\|\boldsymbol{\delta}\|\|_p \leq \epsilon$ |
| Open-set | Grid-level | $\max_{\boldsymbol{\delta}} \min\{d(S(\boldsymbol{x_i} + M \cdot \mathcal{T}(\boldsymbol{\delta})), S(\boldsymbol{x}_i^*))\}_{i=1}^{N}$ |

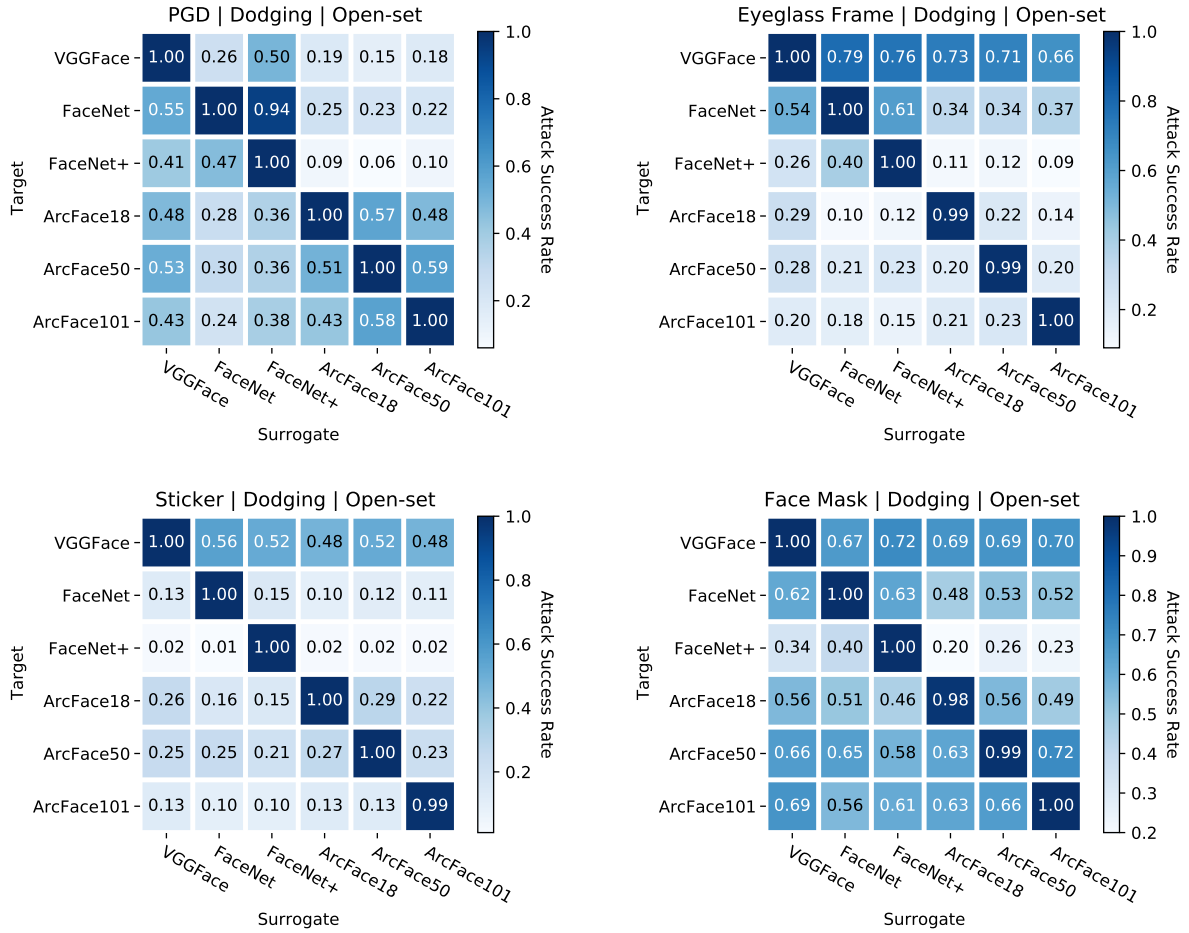


Figure 2. Attack success rate of dodging attacks with different open-set targets and surrogate models. Upper left: PGD attack. Upper right: Eyeglass frame attack. Lower left: Sticker attack. Lower right: Face mask attack.

ilar to those observed from dodging attacks on closed-set systems in the main paper. First, in most cases, an open-set system's neural architecture is more fragile than its training set. For example, under the PGD attack, adversarial examples in response to FaceNet+ have a $94\%$ success rate on FaceNet (which is trained using the same architecture but different training data), while the success rates among the ArcFace systems (which are built with the same training set but different neural architectures) are only around $50\%$. However, there are also some cases where the neural architecture exhibits similar robustness to the training set. For example, when black-box attacks are too weak (under sticker attack), both neural architecture and training set are robust; when the attacks are too strong (under face mask attack), these two components exhibit similar levels of vulnerability. Second, the grid-level face mask attack is considerably more effective than the PGD attack, and significantly more potent than other physically realizable attacks. Like dodging attacks in closed-set settings, most black-box pixel-level physically realizable attacks have relatively low transferability on open-set face recognition systems, with only about $20\%$ success rate.

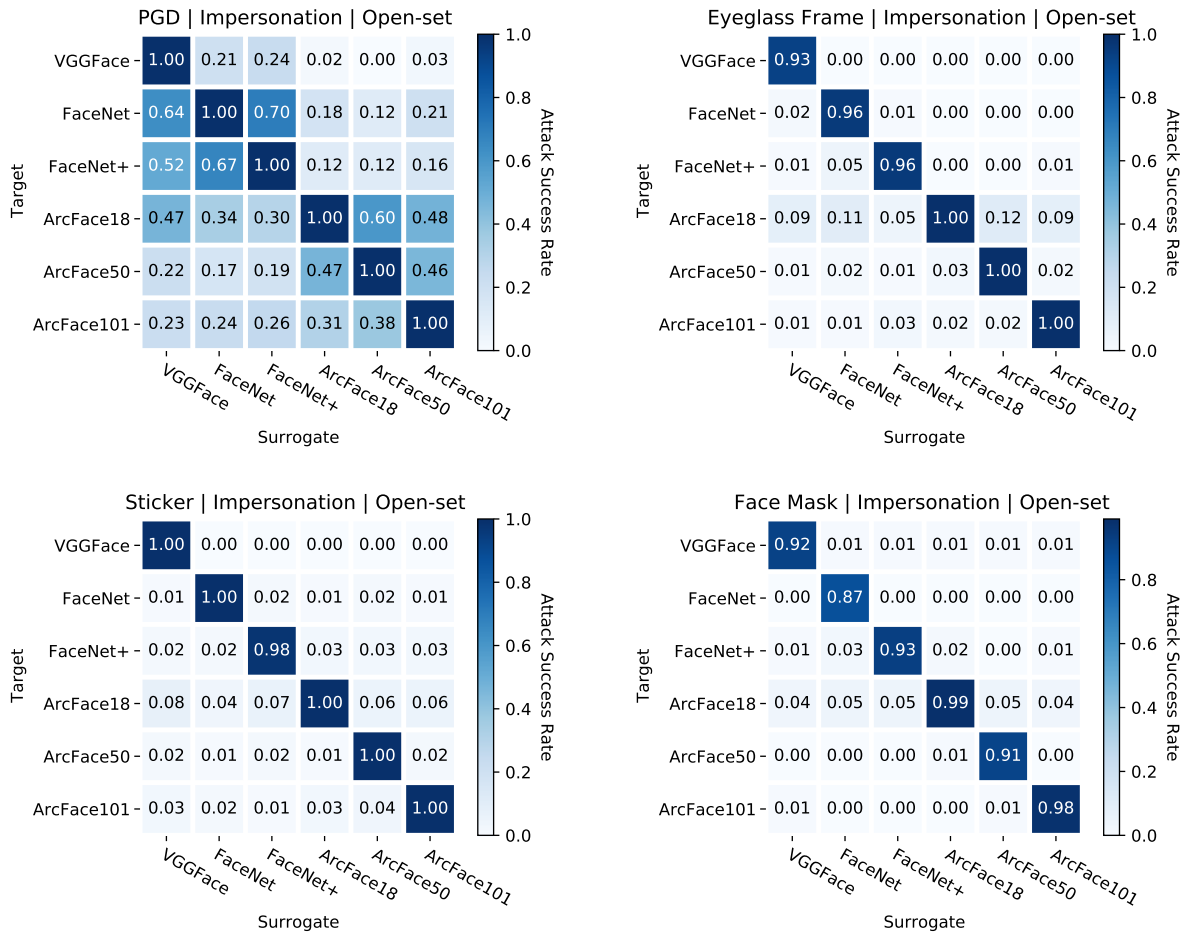


Figure 3. Attack success rate of impersonation attacks with different open-set targets and surrogate models. Upper left: PGD attack. Upper right: Eyeglass frame attack. Lower left: Sticker attack. Lower right:Face mask attack.

## C.2. Closed-set Systems Under Impersonation Attacks

Here, we use impersonation attacks to evaluate the robustness of closed-set systems. In our experiments, all the closed-set models are 100-class classifiers, as introduced in Section 4.1 of the main paper. For any input face image $\boldsymbol{x}$ and its identity $y \in [0, 99]$, we let the target identity of the impersonation attack to be $y_t = (y+1)\%100$. An impersonation attack is successful only when the resulting adversarial example is misclassified as the target identity $y_t$. The results are shown in Table 3.

We have two key findings. First, compared to Table 3 of the main paper, we observe that closed-set systems are significantly more robust to impersonation attacks than dodging attacks. Especially when an attacker has no accurate knowledge about the target system, the attack success rate of physically realizable attacks can be as low as $0\%$. Second, it can be seen that closed-set systems exhibit moderate robustness against digital impersonation attacks. In such attacks, the knowledge of neural architecture is significantly more important than the training set. For example, by knowing the neural architecture of ArcFace18, a PGD attack can achieve a $69\%$ success rate. In contrast, this rate drops to $25\%$ when only the training set is visible to the attacker.

## C.3. Open-set Systems Under Impersonation Attacks

To evaluate impersonation attacks on open-set systems, we randomly select 100 pairs from the LFW dataset [2] in a way similar to Section 4.1 of the main paper. Each pair contains two face images corresponding to different identities. We let one image as the input $\boldsymbol{x}$ and the other as the target gallery image $\boldsymbol{x}_t^*$. An impersonation attack is successful only when the resulting adversarial example and $\boldsymbol{x}_t^*$ are verified as the same identity. The experimental results are presented in Fig. 3.

Similar to the impersonation attacks on closed-set systems, we have the following observations that are consistent

Table 3. Attack success rate of impersonation attacks on closed-set face recognition systems by the attacker's system knowledge. Z represents zero knowledge, T is training set, A is neural architecture, and F represents full knowledge.

| Target System | Attack Type | Attacker's System Knowledge | | | |
|---|---|---|---|---|---|
| | | Z | T | A | F |
| VGGFace | PGD | 0.11 | 0.21 | 0.35 | 1.00 |
| | Eyeglass Frame | 0.01 | 0.01 | 0.03 | 0.95 |
| | Sticker | 0.00 | 0.00 | 0.00 | 1.00 |
| | Face Mask | 0.00 | 0.01 | 0.02 | 1.00 |
| FaceNet | PGD | 0.23 | 0.32 | 1.00 | 1.00 |
| | Eyeglass Frame | 0.00 | 0.00 | 0.28 | 0.99 |
| | Sticker | 0.01 | 0.00 | 0.21 | 1.00 |
| | Face Mask | 0.00 | 0.00 | 0.26 | 0.99 |
| ArcFace18 | PGD | 0.18 | 0.25 | 0.69 | 1.00 |
| | Eyeglass Frame | 0.01 | 0.01 | 0.05 | 0.89 |
| | Sticker | 0.00 | 0.00 | 0.01 | 0.94 |
| | Face Mask | 0.01 | 0.01 | 0.03 | 0.77 |
| ArcFace50 | PGD | 0.13 | 0.15 | 0.45 | 0.87 |
| | Eyeglass Frame | 0.02 | 0.02 | 0.03 | 0.67 |
| | Sticker | 0.00 | 0.00 | 0.00 | 0.58 |
| | Face Mask | 0.01 | 0.01 | 0.01 | 0.60 |
| ArcFace101 | PGD | 0.14 | 0.16 | 0.42 | 0.96 |
| | Eyeglass Frame | 0.00 | 0.00 | 0.03 | 0.58 |
| | Sticker | 0.00 | 0.00 | 0.00 | 0.50 |
| | Face Mask | 0.01 | 0.01 | 0.04 | 0.73 |

Table 4. Attack success rate of dodging PGD attacks on closed-set face recognition systems. Here, only the target system's training data is visible to the attacker, and we use different surrogate models.

| Target System | Surrogate System | | | |
|---|---|---|---|---|
| | Single | | Ensemble | |
| | w/o mmt | w/ mmt | w/o mmt | w/ mmt |
| VGGFace | 0.08 | 0.16 | 0.43 | 0.51 |
| FaceNet | 0.42 | 0.52 | 0.73 | 0.83 |
| ArcFace18 | 0.42 | 0.51 | 0.87 | 0.92 |
| ArcFace50 | 0.35 | 0.55 | 0.86 | 0.90 |
| ArcFace101 | 0.32 | 0.39 | 0.71 | 0.78 |

with our previous summary. First, open-set systems are very robust to black-box impersonation physically realizable attacks. In most cases, these attacks can only achieve a success rate of less than $10\%$. In contrast, the PGD attack is significantly more potent. And under this attack, the neural architecture is considerably more vulnerable than the training set (*e.g.*, comparing FaceNet variations to ArcFace models).

## D. Efficacy of Momentum and Ensemble Models in Transfer-based Attacks

Next, we evaluate the efficacy of using momentum and ensemble-based surrogate models in transfer-based dodging attacks. For a given closed-set target face recognition system, we first train a surrogate model using the same training data. Specifically, we use both a *single* surrogate trained on a different architecture[1], and an *ensembled* surrogate by ensembling the other four systems in the way described in Section 3.2 of the main paper. We then produce white-box dodging attacks on the surrogate and evaluate the resulting examples' attack success rate on the target model. For each attack, we compare the momentum method (*i.e.*, w/ mmt) and the conventional gradient-based approach (*i.e.*, w/o mmt). The results are shown in Table 4, 5, 6, and 7.

Table 5. Attack success rate of dodging eyeglass frame attacks on closed-set face recognition systems. Here, only the target system's training data is visible to the attacker, and we use different surrogate models.

| Target System | Surrogate System | | | |
|---|---|---|---|---|
| | Single | | Ensemble | |
| | w/o mmt | w/ mmt | w/o mmt | w/ mmt |
| VGGFace | 0.17 | 0.22 | 0.26 | 0.28 |
| FaceNet | 0.08 | 0.09 | 0.14 | 0.16 |
| ArcFace18 | 0.02 | 0.03 | 0.05 | 0.06 |
| ArcFace50 | 0.05 | 0.05 | 0.10 | 0.12 |
| ArcFace101 | 0.02 | 0.03 | 0.02 | 0.03 |

Table 6. Attack success rate of dodging sticker attacks on closed-set face recognition systems. Here, only the target system's training data is visible to the attacker, and we use different surrogate models.

| Target System | Surrogate System | | | |
|---|---|---|---|---|
| | Single | | Ensemble | |
| | w/o mmt | w/ mmt | w/o mmt | w/ mmt |
| VGGFace | 0.02 | 0.02 | 0.06 | 0.06 |
| FaceNet | 0.00 | 0.00 | 0.01 | 0.01 |
| ArcFace18 | 0.00 | 0.00 | 0.01 | 0.01 |
| ArcFace50 | 0.00 | 0.00 | 0.00 | 0.01 |
| ArcFace101 | 0.00 | 0.01 | 0.04 | 0.04 |

Table 7. Attack success rate of dodging face mask attacks on closed-set face recognition systems. Here, only the target system's training data is visible to the attacker, and we use different surrogate models.

| Target System | Surrogate System | | | |
|---|---|---|---|---|
| | Single | | Ensemble | |
| | w/o mmt | w/ mmt | w/o mmt | w/ mmt |
| VGGFace | 0.18 | 0.26 | 0.20 | 0.32 |
| FaceNet | 0.26 | 0.38 | 0.42 | 0.42 |
| ArcFace18 | 0.21 | 0.33 | 0.21 | 0.33 |
| ArcFace50 | 0.28 | 0.34 | 0.36 | 0.36 |
| ArcFace101 | 0.22 | 0.34 | 0.30 | 0.36 |

We have two key observations. First, both ensemble and momentum contribute to stronger transferability, although in most cases, ensemble contributes more. For example, the ensemble method can boost the transferability of PGD attacks on FaceNet by $31\%$, while the improvement by momentum is only about $10\%$. Second, the efficacy of momentum and ensemble models is highly dependent on the nature of perturbation. For digital attacks, these methods combined can significantly improve transferability by up to $55\%$. In

[1]For a given target model, we trained four single surrogates corresponding to the other four architectures. Below, we only present the result of the surrogate that has the highest attack success rate.

Table 8. Attack success rate of dodging attacks on open-set face recognition systems by the universality of adversarial examples. Here, $N$ represents the batch size of face images that share a universal perturbation.

| Target System | Attack Type | Attacker's Capability | | | |
|---|---|---|---|---|---|
| | | N=1 | N=5 | N=10 | N=20 |
| VGGFace | PGD | 1.00 | 0.89 | 0.81 | 0.53 |
| | Eyeglass Frame | 1.00 | 1.00 | 1.00 | 1.00 |
| | Sticker | 1.00 | 1.00 | 1.00 | 1.00 |
| | Face Mask | 1.00 | 1.00 | 1.00 | 1.00 |
| FaceNet | PGD | 1.00 | 0.02 | 0.02 | 0.02 |
| | Eyeglass Frame | 1.00 | 1.00 | 1.00 | 1.00 |
| | Sticker | 1.00 | 1.00 | 0.99 | 0.90 |
| | Face Mask | 1.00 | 1.00 | 0.99 | 0.98 |
| ArcFace18 | PGD | 1.00 | 0.96 | 0.79 | 0.46 |
| | Eyeglass Frame | 0.99 | 0.86 | 0.70 | 0.67 |
| | Sticker | 1.00 | 1.00 | 1.00 | 0.99 |
| | Face Mask | 0.98 | 0.98 | 0.93 | 0.92 |
| ArcFace50 | PGD | 1.00 | 0.91 | 0.75 | 0.47 |
| | Eyeglass Frame | 0.99 | 0.78 | 0.67 | 0.62 |
| | Sticker | 1.00 | 1.00 | 1.00 | 0.00 |
| | Face Mask | 0.99 | 0.99 | 0.99 | 0.94 |
| ArcFace101 | PGD | 1.00 | 0.68 | 0.68 | 0.41 |
| | Eyeglass Frame | 1.00 | 0.85 | 0.73 | 0.65 |
| | Sticker | 0.99 | 0.98 | 0.97 | 0.97 |
| | Face Mask | 1.00 | 1.00 | 1.00 | 1.00 |

grid-level face mask attacks, the improvement is as considerable as up to $16\%$. However, both methods can only marginally boost the transferability of pixel-level realizable attacks. Especially in the sticker attacks, the improvement is nearly negligible. We leave effective transfer-based pixel-level physically realizable attacks as an open problem for future research.

## E. Universal Attacks

Finally, we evaluate open-set systems under universal dodging attacks. The results are shown in Table 8. Compared to Table 5 of the main paper, we find that open-set systems are significantly more fragile to universal perturbations of all types than their closed-set counterparts. For example, when $N = 20$, the open-set ArcFace101 is susceptible to all the four types of universal attacks, while in the closed-set setting it is only vulnerable to the universal face mask attack. Moreover, we again observe that the universal grid-level face mask attack is more effective than the other perturbation types. Here, we also find that the sticker attack is as potent as the face mask attack in open-set settings.

## References


[1] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[2] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[3] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. Universal adversarial perturbations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1765–1773, 2017.

[4] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.