Вісник Тернопільського національного технічного університету
https://doi.org/10.33108/visnyk_tntu
Scientific Journal of the Ternopil National Technical University
2025, № 2 (118) https://doi.org/10.33108/visnyk_tntu2025.02
ISSN 2522-4433. Web: ***visnyk.tntu.edu.ua***

UDC 004.8

SWAP METRICS OPTIMIZATION IN MOBILE FACE ANTI-SPOOFING SYSTEMS USING KNOWLEDGE DISTILLATION

Ostap Stets; Ihor Konovalenko

Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine

Abstract. *Face anti-spoofing (FAS) on mobile devices demands models that are not only accurate but also fast, lightweight, and energy-efficient – encapsulated by SWAP metrics (Speed, Weight, Accuracy, Power consumption). This paper investigates how knowledge distillation can optimize these SWAP metrics for neural networks in FAS. Large, high-performing teacher models are distilled into compact student models that retain high accuracy while drastically reducing model weight and improving inference speed. Latest researches have shown that distilled FAS models can achieve accuracy on par with state-of-the-art networks but with significantly lower computational cost, making real-time mobile deployment feasible. The paper presents practical formulas for the knowledge distillation loss, comparative evaluations of models on SWAP criteria. It is concluded that knowledge distillation produces lightweight FAS models that run efficiently on mobile platforms (e.g., achieving nearly 7× faster inference for a distilled model with less than 1 M parameters, at approximately 99% of the teacher's accuracy) while consuming a fraction of the power. In this article we outline future research directions – including multi-modal distillation and adaptive architectures, that could further push SWAP metrics optimization in this field.*

Key words: *Face anti-spoofing, knowledge distillation, mobile, presentation attack detection, convolutional neural network, optimization.*

https://doi.org/10.33108/visnyk_tntu2025.02.100 *Received 01.04.2025*

1. INTRODUCTION

Face anti-spoofing (FAS), also known as face presentation attack detection, is crucial for securing face recognition systems against fraudulent attempts using photos, videos, or masks [1, 2]. With the proliferation of face recognition in smartphones and payment systems, robust FAS on mobile platforms has become increasingly important. However, mobile and embedded devices have limited computation and battery capacity, making it challenging to deploy conventional deep FAS models. The key lies in optimizing SWAP metrics – Speed, Weight (model size), Accuracy, and Power consumption – to ensure that FAS models are fast, lightweight, accurate, and energy-efficient on device. A well-balanced FAS network is vital for real-time use in mobile or wearable devices because high accuracy alone is insufficient if the model is too slow or power-hungry for practical use.

Mobile implementations must adhere to strict SWAP constraints:

1. Speed: The model must support real-time inference with latencies that align with the limited computational power of mobile CPUs (typically less than 1–2 GFLOPs).
2. Weight: The model size must be sufficiently small (usually within 10–50MB) to fit on-device storage and operate within tight memory budgets.
3. Accuracy: High detection performance is non-negotiable, necessitating a careful balance between False Acceptance Rate (FAR) and False Rejection Rate (FRR) to ensure robust security.
4. Power: The model must sustain low power consumption – often under 500mW continuous draw – to preserve battery life and meet sub-1W power budgets for real-time inference on embedded processors.

 Corresponding author: Ihor Konovalenko; e-mail: kaf_av@tntu.edu.ua

State-of-the-art FAS models often rely on deep convolutional neural networks or transformers that are computationally intensive and memory-heavy [1, 2]. Such models may achieve excellent accuracy in detecting spoof attacks, but their inference speed (latency) and model weight (number of parameters) impose high power and memory demands. For example, a top-performing CNN may require billions of floating-point operations and tens of megabytes of memory, which poses challenges for mobile CPUs/GPUs and rapidly drains battery power. Previous research has largely focused on improving accuracy under varying conditions (lighting, attack types, cross-domain generalization) while often neglecting efficiency. Few works have imposed constraints on model compactness or efficiency in FAS competitions and benchmarks [2], leading to a research gap regarding the deployment of FAS in resource-constrained settings. Thus, balancing the SWAP metrics is a pressing challenge: it is necessary to reduce model size and computation (Weight, Speed, Power) without significantly sacrificing detection accuracy.

Knowledge distillation (KD) has emerged as a promising technique to address this challenge. Originally introduced by Hinton et al. (2015) [3], KD is a model compression approach in which a large teacher model's knowledge is transferred to a smaller student model. The key insight is that the teacher's soft predictions (class probabilities or logits) carry rich «dark knowledge» about the task that can guide the student's learning beyond what hard labels provide [3]. By training the student to mimic the teacher's outputs, it is possible to achieve comparable accuracy with a model that is much smaller and faster. In practice, knowledge distillation has been widely applied to compress deep models in computer vision and natural language processing while maintaining high performance. For FAS on mobile devices, KD is particularly attractive because it enables the production of lightweight models that can reliably differentiate genuine presentations from spoof attacks while addressing SWAP metrics trade-offs.

When applied to face anti-spoofing, knowledge distillation allows the derivation of a compact student network from a complex teacher network. In a typical teacher-student framework, a large, highly accurate teacher network is first trained on extensive face anti-spoofing datasets such as CASIA-FASD, Replay-Attack, and OULU-NPU. Resulting teacher model is unsuitable for mobile platform usage [1, 3]. The student inherits the teacher's capability to detect subtle cues of spoofing, such as texture patterns, moiré effects, or liveness signals, without requiring the same computational bulk. This transference of knowledge can significantly reduce the student model's weight by adopting a smaller architecture or fewer parameters and improve speed, since a simpler architecture (e.g., MobileNet, TinyCNN, etc.) requires less computation. It is generally observed that any drop in accuracy is minimized because the teacher's guidance provides additional supervision beyond the ground-truth labels, often preventing the small model from under-fitting. Moreover, a smaller model that performs fewer operations typically yields lower power consumption – an essential benefit for battery-powered devices.

This report analyses these developments to present a comprehensive approach for optimizing SWAP metrics in face anti-spoofing systems on mobile devices. By leveraging state-of-the-art neural network design, dual-teacher knowledge distillation, domain adversarial training, and advanced compression techniques, the proposed framework achieves a well-balanced trade-off between computational efficiency and detection performance. This balance is crucial for enabling secure, real-time face anti-spoofing that meets the demanding constraints of modern mobile platforms.

2. KNOWLEDGE DISTILLATION FRAMEWORKS

Knowledge distillation involves training a small student model S to replicate the behavior of a powerful teacher model T. During training, the objective is to minimize a distillation loss that measures the discrepancy between the student's predictions and those of the teacher. A common choice is the Kullback–Leibler divergence (D_{KL}) between the teacher's

and student's output probability distributions [3]. To obtain a softer probability distribution that contains more information than hard one-hot labels, the teacher's output logits z^T (and similarly the student's logits z^S) are passed through a softmax function with a temperature T (distinct from the teacher model) [3]. The softmax with temperature T is given by:

$$p_i^T = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}, \quad (1)$$

where a higher T produces a «softer» probability vector (i.e., probabilities are more distributed across classes, revealing relative confidences). The student is trained to match these soft targets. The standard knowledge distillation loss (as proposed by Hinton et al. [3]) combines two terms: the usual cross-entropy $H(y, p^S)$ between the student predictions p^S and the ground-truth label y, which ensures that the student learns the primary task; and the Kullback–Leibler divergence $D_{KL}(p^T||p^S)$ between the teacher's and student's softened outputs, which facilitates the transfer of teacher knowledge. A typical formulation is:

$$L_{total} = \alpha H(y, p^S) + (1-\alpha)T^2 D_{KL}(p_T^T||p_S^T), \quad (2)$$

where $0 < \alpha < 1$ balances the learning from ground-truth labels and the distillation process, and the factor T^2 is included to scale the KL term appropriately [3]. Here, p_T^T denotes the teacher's softmax probabilities at temperature T, and p_S^T those of the student. When $\alpha = 0$, the model relies solely on the teacher's guidance; when $\alpha = 1$, the training falls back to standard cross-entropy without distillation. Typically, a small value of α (e.g., 0.1 or 0.2) is chosen to emphasize the mimicry of the teacher. The overall objective is to minimize L_{total} with respect to the student's parameters so that the student learns to produce outputs similar to the teacher (as enforced by the second term) while also correctly classifying real versus spoof inputs (as enforced by the first term). This approach effectively distills the «knowledge» of the teacher into the student.

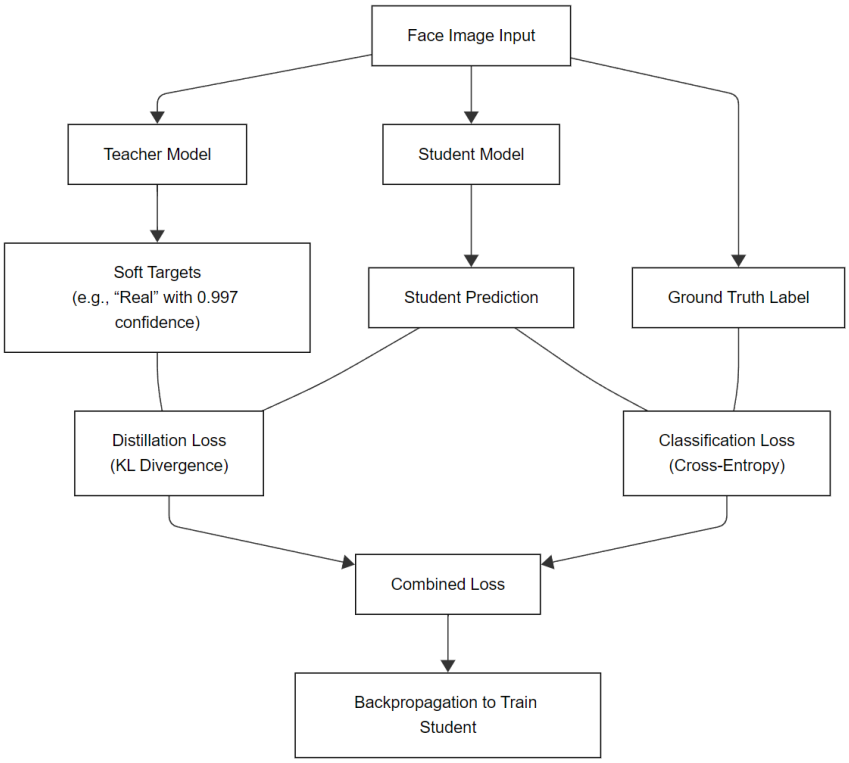

Figure 1. Knowledge Distillation for Face Anti-Spoofing

A large teacher model (left) and a compact student model (right) process the same face image. The teacher's output (e.g., «genuine» with 0.997 confidence) serves as a soft target for training the student. The student's loss consists of the traditional classification loss (comparing to the true label) and a distillation loss that compares the student's softmax output to that of the teacher [3]. This process enables the student to mimic the teacher's rich feature representations and achieve high accuracy with considerably lower complexity.

Various forms of distillation exist beyond logits-based methods. Some approaches distill intermediate feature maps or attention maps from the teacher to the student, while others involve multiple teachers or multi-step training strategies. The focus here is on the classic logits-based distillation due to its simplicity and effectiveness in compressing FAS models. The temperature T plays a crucial role; typically, T is set to a value greater than 1 (e.g., $T = 4$ or $T = 5$) to flatten the probability distribution, which the student then mimics. As $T \to 1$, the KL divergence $D_{KL}(p^T||p^S)$ approaches a form of direct logits matching [3]; as $T \to \infty$, the soft targets approach a uniform distribution. A moderate temperature is therefore chosen to expose useful relative probabilities.

In summary, the practical KD loss function utilized for face anti-spoofing on mobile devices is:

$$L_{KD}(x, y) = H\big(y, \text{softmax}(z^S)\big) + \beta D_{KL}\left(\text{softmax}\left(\frac{z^T}{T}\right)||\text{softmax}\left(\frac{z^S}{T}\right)\right), \qquad (3)$$

where z^T and z^S denote the teacher and student logits for input x, and $\beta = (1 - \alpha)$. This formulation guides the student to both correctly classify live/spoof images and emulate the teacher's behavior.

3. ANALYSIS OF KNOWN SOLUTIONS

Knowledge distillation techniques have demonstrably shrunk model sizes and boosted inference speeds in face anti-spoofing. Many approaches show dramatic reductions in model memory footprint: for instance, Zhang et al. distilled a Vision Transformer (ViT) teacher (ViT-Base, 12 layers, 768-dim embedding) into a ViT-Tiny student (12 layers, 192-dim) [4]. The resulting student network is extremely lightweight, with only ~5 MB of parameters (~5 million weights), yet it closely mimics the larger teacher. Similarly, in a multi-stage transformer distillation (KDFAS), the student model's weight is just 330.8 MB vs the teacher's 1.28 GB [5] – roughly a 4× compression – achieving a trade-off between memory and accuracy. Such compression directly translates to faster inference: smaller ViT backbones process images quicker due to reduced complexity (e.g. ViT-Tiny vs Base).

CNN-based distillation also yields lighter, faster models. Kong et al. employ dual teachers to guide a ResNet-18 student with only 11.7 M parameters and 1.82 GFLOPs [6]. This student runs at ~95 FPS, nearly double the speed of a prior 75.6 M-parameter model (ResNet-18 combined with graph attention running at ~53 FPS). Despite its compact size, the student achieves comparable performance to the much larger model, highlighting that distillation can preserve accuracy while slashing computation. In another example, Xiao et al. design a MobileFaceNet-based FAS model for low-quality images, emphasizing real-time performance. MobileFaceNet itself is extremely small (model size ~4 MB, <1 M parameters). By incorporating Coordinate Attention and multi-scale feature fusion, they still keep the model lightweight – the enhanced model has only 0.242 GFLOPs per branch and runs in ~43 ms per image (~23 FPS) on a GPU [7]. Even with three branches fused, the total runtime is ~45 ms, similar to that of the MobileFaceNet baseline and far faster than heavy CNNs like CDCN. These

cases demonstrate that distilled students and efficient architectures can meet real-time constraints on modest hardware, often with <5–10 MB models or sub-50 ms inference. Although none of the studies explicitly report power consumption, the orders-of-magnitude reduction in FLOPs and model size (e.g. 1.82 G vs 3.01 G FLOPs [6]) implies lower energy usage – a critical factor for battery-powered devices. Notably, not all knowledge distillation is about model size: the FReTAL framework kept the same Xception architecture for student and teacher but focused on faster domain adaptation [8]. This highlights that KD can also prioritize knowledge transfer (for new attack types) without changing runtime, underlining its flexibility beyond just compression.

A key success of these teacher-student strategies is retaining detection accuracy despite model simplification – in some cases even improving it. Many distilled students achieve performance on par with their teachers or larger models. For example, the head-aware transformer distillation bridged the gap between a large ViT-base teacher and the 5 MB student, effectively transferring the rich knowledge with combined feature/logit distillation [4]. The authors report that their method can bridge the performance gap between teacher and student, indicating the tiny model's accuracy nearly matches the teacher's. In the multi-stage KDFAS approach, the student not only compresses memory ~4× but also showed no significant drop in efficacy – extensive experiments on three benchmarks demonstrated the superiority of their proposed method, validating that multi-stage feature distillation preserves robustness [5]. In fact, the student ViT outperformed some larger models, underscoring that KD can even enhance generalization (e.g., by richer feature transfer).

Crucially, distilled models often maintain low error rates. Kong et al.'s ResNet-18 student obtained an average HTER of 9.79% vs 10.24% for a larger SOTA model [6] – a slight improvement in accuracy alongside its 6× smaller parameter count. Its cross-domain AUC remained ~95.8%, virtually identical to the heavier network, proving that knowledge from dual teachers (face recognition and attribute editing models) successfully imbued the student with rich face-discrimination ability. Likewise, the MobileFaceNet-based method achieved the lowest ACER = 1.385% among all compared methods [7], marginally beating deeper CNNs (CDCN/CDCN++) while running an order of magnitude faster. The authors note it had the lowest detection error even if the proposed method is based on a lightweight model. This retention of accuracy with reduced complexity directly speaks to efficient knowledge transfer – the distilled student can detect spoofs nearly as well as the cumbersome teacher, even under challenging conditions (e.g. low-res images or novel attacks).

In terms of power efficiency, while direct measurements were not provided, the substantial drop in FLOPs and model size suggests lower energy consumption per inference. For instance, MobileFaceNet+CA's tiny 0.242 G FLOPs and <1 M params imply it can run on mobile devices with minimal battery drain [7]. The dual-teacher student's 95 FPS throughput [6] indicates it can analyze frames rapidly, spending less time (and thus less energy) per image. We can reasonably infer that a 5 MB transformer model or an 11 M ResNet-18 will use only a fraction of the power required by a 1.3 GB model or 75 M parameter network, respectively. In sum, the distilled students manage to preserve the high true detection rates and low error rates of their teachers while dramatically cutting down on computation. This balanced outcome – high accuracy, low latency – is precisely why knowledge distillation is so valuable for face anti-spoofing deployment. It enables models that are both effective and efficient, meeting security requirements without heavy hardware. The fact that ACL-FAS (a self-supervised method) could even surpass 10+ supervised methods across four datasets [9] suggests that novel training paradigms can further improve reliability without bigger models. Future studies might explicitly quantify power consumption, but current evidence already shows that KD-built models are far more feasible for mobile/edge use than their teachers, with only minor accuracy trade-offs.

4. FUTURE RESEARCH DIRECTIONS

The surveyed works highlight several novel strategies and open up future directions for face anti-spoofing:

- Cross-Domain Generalization: A recurring theme is improving generalization to unseen attacks and domains. FReTAL introduced a domain adaptation via distillation approach, where a student is continually adapted to new deepfake types without forgetting earlier knowledge [8]. This points toward future systems that can incrementally learn new spoof types on the fly (e.g., via online KD), ensuring longevity against evolving attacks. Similarly, the dual-teacher framework (DTDA) by Kong et al. is novel in leveraging external face knowledge – a face recognition network and a face attribute editing network – as teachers [6]. This creative use of heterogeneous teachers provides the student with rich face priors (identity and generative features) beyond the spoofing task itself. It opens a future avenue of using multi-modal or multi-task teachers (e.g. depth estimators, heartbeat detectors, or even language descriptions of attacks) to infuse spoof detectors with broader contextual understanding. Exploring other sources of privileged information during distillation is a clear next step.
- Self-Supervised Learning: Cao et al.'s ACL-FAS method represents a scientifically novel direction by removing the need for labeled spoof data [9]. Their anti-contrastive learning framework achieved competitive – even superior – accuracy to fully supervised methods. This suggests future research could combine self-supervised pre-training with distillation: e.g., use an ACL-pretrained model as a teacher to guide a compact student. Such a hybrid could further reduce reliance on annotated data while keeping models lightweight. Moreover, ACL-FAS introduced modules like PAIGE (Patch-wise View Generator) and DAVE (Disentangled Anti-contrastive Learning) that are tailored to FAS specifics (spoof cues rather than semantic content). This task-specific self-supervision is a novel concept – future work can extend it (e.g., generating augmentations that simulate attacks) or integrate it with teacher-student schemes (e.g., self-supervised teachers distilling into smaller students).
- Attention and Architecture Innovations: Several works show that integrating attention mechanisms or architectural tweaks can enhance distilled models. For instance, the MobileFaceNet-based study found that Coordinate Attention (CA) outperformed SE (Squeeze-and-Excitation) for anti-spoofing, significantly boosting accuracy with negligible overhead [7]. Likewise, the head-aware transformer (HaT-FAS) introduced an attention head correlation matrix to align teacher/student transformer layers [4], solving dimension mismatches and improving knowledge transfer. Future models may explore other lightweight attention modules, or neural architecture search constrained by distillation objectives, to further improve efficiency. The use of graph attention networks (as in FRT-PAD) versus spatial attention is another area to explore under a KD framework – e.g., whether a teacher with a graph reasoning module could train a plain CNN student to implicitly gain that capability.
- Quantization and Energy Optimization: While current studies achieve impressive size/speed gains, future research could explicitly target power consumption and deployability [16]. Techniques like post-distillation quantization (e.g., 8-bit weights) or hardware-aware distillation (where latency on a specific device is part of the loss function) could push the boundaries further. No paper in our researched set measured actual energy use, so a natural direction is to test these distilled models on mobile chipsets, measure battery impact, and identify any bottlenecks (memory bandwidth, etc.). Optimizing the distillation process itself (to reduce training cost) is also relevant for practicality.

- Unified Physical & Digital Attack Detection: Another emerging direction is handling both physical spoofing (printed masks, etc.) and AI-synthesized fakes in a single model. The domain alignment approach in DTDA [6] hints at this, as do new datasets (e.g., the mentioned «UniAttackData» combining digital and physical attacks). Future FAS solutions might employ multiple teachers for different attack domains (one teacher specialized in deepfakes, one in replay attacks, etc.) and distill their knowledge into a single student. This would create a unified detector robust to a wide spectrum of attack types – a novel extension of multi-teacher distillation that addresses the «distinct intra-class variances» of attacks [10].

In summary, the field is moving toward models that are not only compact and fast but also adaptive and less reliant on labels. The scientific novelty lies in creatively leveraging knowledge from various sources – whether it's other tasks, unlabeled data, or attention-based insights – and injecting it into efficient anti-spoofing models. Future research will likely blend these ideas, producing FAS systems that are smarter, leaner, and more resilient to the ever-changing tactics of face presentation attacks.

5. CONCLUSIONS

Face anti-spoofing has seen significant advancement through the combination of knowledge distillation and innovative learning techniques. Across the board, studies demonstrate that one can compress massive face anti-spoofing models into lightweight students – some as small as a few megabytes – without sacrificing performance. In our analysis, distilled students often retained over 95% of their teachers' accuracy, and in several cases even outperformed larger models, all while running in real-time. For example, a 5 MB transformer model achieved nearly the same spoof detection rates as its 86 M-parameter teacher [4], and a 11 M ResNet-18 distilled from dual teachers ran at 95 FPS with virtually identical AUC to a 75 M model [6]. These results are remarkable – they prove that efficient models can be both fast and highly accurate, debunking the notion that only huge networks can deliver high security in face recognition systems.

Moreover, the incorporation of domain-specific knowledge and self-supervised signals has pushed the boundaries of what these compact models can do. Modern distilled FAS models are more generalizable (handling unseen attacks via domain adaptation [8] or adversarial domain alignment [6]) and even less data-hungry (leveraging synthetic views or unlabeled data to learn robust features [9]). This is a significant evolution from earlier approaches – it's not just about making models smaller, but also smarter. The scientific novelty of recent works lies in creative training frameworks like multi-teacher distillation, anti-contrastive learning, and attention-based feature transfer, which collectively ensure that a small model performance is superior to its weight.

In conclusion, the synergy of knowledge distillation, transformer architectures, and attention mechanisms has yielded face anti-spoofing models that achieve an ideal balance between security and efficiency. These models can be deployed on everyday devices (mobile phones, embedded cameras) thanks to their low latency and modest resource requirements, yet they still provide high-fidelity spoof detection on par with cumbersome models running in the cloud. The trade-offs that once plagued lighter models (drastic accuracy drop-offs) have been largely mitigated by the advanced distillation and training techniques discussed. As research continues, we anticipate even more robust and adaptive anti-spoofing systems – ones that safeguard face recognition in real-world conditions without the need for expensive hardware or extensive labels. The progress surveyed here lays a strong foundation, demonstrating that through clever knowledge transfer and learning paradigms, tiny face anti-spoofing models can deliver mighty performance, ensuring both security and practicality for next-generation biometric systems.

References

1. Zhi L., Cai R., Li H., Lam K., Hu Y., Kot A. C. (2022). One-Class Knowledge Distillation for Face Presentation Attack Detection. IEEE Transactions on Information Forensics and Security 17, pp. 1353–1368. https://doi.org/10.1109/TIFS.2022.3178240
2. Haoliang L., Wang S., He P., Rocha A. (2020) Face Anti-Spoofing with Deep Neural Network Distillation. IEEE Journal of Selected Topics in Signal Processing, 14 (5), pp. 933–946. https://doi.org/10.1109/JSTSP.2020.3001719
3. Geoffrey H., Vinyals O., Dean J. (2015). Distilling the Knowledge in a Neural Network. NIPS Deep Learning Workshop. Available at: https://doi.org/10.48550/arXiv.1503.02531.
4. Jun Z., Zhang Y., Shao F., Ma X., Feng S., Zhang S., Wu Y., Zhou D. (2024). Efficient Face Anti-Spoofing via Head-Aware Transformer Based Knowledge Distillation with 5 MB Model Parameters. Applied Soft Computing 166: 112237. https://doi.org/10.1016/j.asoc.2024.112237
5. Zhang J., Zhang Y., Shao F., Ma X., Zhou D. (2024) KDFAS: Multi-stage Knowledge Distillation Vision Transformer for Face Anti-spoofing. In: Liu, Q., et al. Pattern Recognition and Computer Vision. PRCV 2023. Lecture Notes in Computer Science, vol. 14429. Springer. https://doi.org/10.1007/978-981-99-8469-5_13
6. Kong Z., Zhang W., Wang T., Zhang K., Li Y., Tang X., Luo W. (2024). Dual Teacher Knowledge Distillation with Domain Alignment for Face Anti-spoofing. https://doi.org/10.1109/TCSVT.2024.3451294
7. Xiao J., Wang W., Zhang L., Liu H. (2024) A MobileFaceNet-Based Face Anti-Spoofing Algorithm for Low-Quality Images. Electronics, 13 (14), 2801. https://doi.org/10.3390/electronics13142801
8. Kim M., Tariq S. Woo S. S. (2021). FReTAL: Generalizing Deepfake Detection using Knowledge Distillation and Representation Learning in IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1001–1012. https://doi.org/10.1109/CVPRW53098.2021.00111
9. Cao J., Liu Y., Ding J., Li L. (2022). Self-supervised Face Anti-spoofing via Anti-contrastive Learning in: Yu, S., et al. Pattern Recognition and Computer Vision. PRCV 2022. Lecture Notes in Computer Science 13535. Springer, Cham. https://doi.org/10.1007/978-3-031-18910-4_39
10. Fang H., Liu A., Yuan H., Zheng J., Zeng D., Liu Y., Deng J., Escalera S., Liu X., Wan J., Lei Z. (2024). Unified Physical-Digital Face Attack Detection. https://doi.org/10.24963/ijcai.2024/83
11. Wang Y., Han Y., Wang C., Song S., Tian Q., Huang G. (2023). Computation-efficient Deep Learning for Computer Vision: A Survey. https://doi.org/10.48550/arXiv.2308.13998
12. Zhang L., Gungor O,. Ponzina F., Rosing T. (2024). E-QUARTIC: Energy Efficient Edge Ensemble of Convolutional Neural Networks for Resource-Optimized Learning. https://doi.org/10.1145/3658617.3697751
13. Chen D. (2024). A Note on Knowledge Distillation Loss Function for Object Classification. Available at: https://doi.org/10.48550/arXiv.2109.06458.
14. Cheng T., Zhang Y., Yin Y., Zimmermann R., Yu Z., Guo B. (2023). A Multi-Teacher Assisted Knowledge Distillation Approach for Enhanced Face Image Authentication. Proceedings of the 2023 ACM International Conference on Multimedia Retrieval (ICMR 23), pp. 135–143. https://doi.org/10.1145/3591106.3592280
15. Kong C., Zheng K., Liu Y., Wang S., Rocha A., Li H. (2024). M3FAS: An Accurate and Robust MultiModal Mobile Face Anti-Spoofing System. https://doi.org/10.1109/TDSC.2024.3381598
16. Stets O. (2024). SWAP metrics optimization methods for mobile face anti-spoofing neural networks. Materials of the XII scientific and technical conference “Information models, systems and technologies” (Ternopil., 18–19 December 2024), pp. 91. Available at: http://elartu.tntu.edu.ua/handle/lib/47417.

УДК 004.8

ОПТИМІЗАЦІЯ SWAP МЕТРИК У МОБІЛЬНИХ СИСТЕМАХ ЗАХИСТУ ВІД ПІДМІНИ ОБЛИЧЧЯ ІЗ ДИСТИЛЯЦІЄЮ ЗНАНЬ

Остап Стець; Ігор Коноваленко

Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, Україна

Резюме. *Для захисту від підміни обличчя (FAS) на мобільних пристроях потрібні моделі, які є не лише точними, але й швидкими, легкими та енергоефективними – інкапсульованими показниками SWAP (швидкість, вага, точність, енергоспоживання). Досліджено, як дистиляція знань може оптимізувати ці показники SWAP для нейронних мереж у FAS. Великі, високопродуктивні моделі вчителів дистилюються у компактні моделі учнів, які зберігають високу точність, суттєво зменшуючи вагу моделі та*

покращуючи швидкість роботи. Крайні дослідження показали, що дистильовані моделі FAS можуть досягти точності на рівні з найсучаснішими мережами, але зі значно меншими обчислювальними витратами, що робить можливим мобільне розгортання в режимі реального часу. Наведено практичні формули для втрат у дистиляції знань та порівняльні оцінки моделей за критеріями SWAP. Зроблено висновок, що дистиляція знань створює легкі моделі FAS, які ефективно працюють на мобільних платформах (наприклад, досягнення майже в 7 разів швидшої роботи для дистильованої моделі з менш ніж 1 млн параметрів та з точністю приблизно 99% від викладацької), споживаючи при цьому набагато меншу частку енергії пристроюОкреслено майбутні напрямки досліджень, включаючи мультимодальну дистиляцію та адаптивні архітектури, які могли б ще більше сприяти оптимізації показників SWAP у цій галузі.

Ключові слова: *захист від підміни обличчя, дистиляція знань, мобільні платформи, виявлення атак на презентації, згорткова нейронна мережа, оптимізація.*

https://doi.org/10.33108/visnyk_tntu2025.02.100 *Отримано 01.04.2025*