

Facial Landmark Detection: An Attentive Dropout-Based Occlusion-Adaptive Deep Network

Muhammad Sadiq¹[0000–0003–2199–3702], Junwei Liang^{*1}[0000–0003–1999–0254],
Yu Geng¹, Yunsheng Zhang¹[0009–0003–1149–5096], and Lu Chen¹

Shenzhen Institute of Information Technology, Shenzhen, China

Abstract. This study introduces the Attentive Dropout-based Occlusion-adaptive Deep Network (ADODN) for facial landmark detection, integrating modules for geometry awareness, attentive dropout, and low-rank learning to address the challenges posed by occlusions, extreme poses, emotional expressions, and variable illumination in Facial Landmark Detection (FLD). Despite the high accuracy of CNN-based FLD methods, occlusion—due to its unpredictable nature—significantly impairs model performance. ADODN mitigates this by employing geometric correlations and an innovative attentive dropout mechanism that selectively removes the most discriminative features, thereby maintaining classification accuracy even in partially occluded scenarios. Our findings, supported by ablation studies, highlight the effective collaboration between ADODN’s modules, enhancing occlusion handling and localization precision. Contributions of this research include the enhancement of existing models with advanced attention and dropout techniques for accurate occlusion management, the novel application of dropout masks for occlusion modeling in FLD, and a methodological reduction in network parameters that decreases both training time and costs, optimizing ADODN for extensive dataset processing.

Keywords: Attention · Facial Landmark Detection · Occlusion · Dropout mask.

1 Introduction

Facial landmarks, critical for advanced facial analysis, are defined as key points on a face surrounding essential features such as the ears, eyes, nose, mouth, and chin. These landmarks are fundamental to tasks varying in complexity and objectives, leading to an expected increase in research focused on their precise localization in the upcoming decade [1].

Facial Landmark Detection (FLD) is central to a range of facial analysis applications, from action unit detection to 3D face modeling, challenged by occlusions, lighting variations, and other dynamic factors. FLD methodologies fall

^{*} Corresponding Author: jwliang@szit.edu.cn



Fig. 1: Few examples of occlusion from COFW dataset [9] caused by hairs, hands, sports wears, food, etc. Examples show clearly that identification of facial landmark is very challenging in presence of occlusion.

into three main categories: regression-based, template-based, and deep learning (DL) approaches.

Regression-based methods connect facial image attributes with landmark positions, eschewing comprehensive shape models for simultaneous landmark prediction and the integration of shape constraints [2]. Template-based strategies manage facial variation through statistical models and Principal Component Analysis (PCA), creating a mathematical facial structure framework from categorized data [3,4,5,6]. These models, however, struggle with occlusion-related inaccuracies.

DL, particularly through Convolutional Neural Networks (CNNs), has significantly advanced FLD, addressing challenges such as occlusion that impede traditional methods [7]. Recent innovations involve integrating attention mechanisms to improve occlusion handling and feature representation, exemplified by the attentive dropout module incorporating Channel-wise Attention (CA), drop mask, and importance mask to refine localization accuracy [8]. This approach underscores DL’s evolving role in enhancing FLD’s effectiveness against the backdrop of occlusions and spatial distortions.

The short summary of our trifold contributions consists of : i) Improved ADN, ODN, and AODN models by using attention, dropout mask, and importance map to effectively handle occlusion and achieve accurate localization. ii) Introduced the utilization of drop mask and importance map in FLD for occlusion modeling, showcasing the exceptional performance of ADODN on difficult datasets. iii) The proposed methodology decreases the number of network parameters, resulting in significant reductions in both training time and cost. This makes it well-suited for processing massive amounts of data.

The subsequent sections of the paper are structured in a systematic fashion. Section 2 offers essential introduction and contextual information. The specifics of our proposed Attentive Dropout-based Occlusion-adaptive Deep Network (ADODN) model can be found in section 3. The subject of mathematical optimization is addressed in section 4. Exhaustive description of the numerous experiments conducted within our proposed framework is in Section 5. The conclusive analysis and findings of this study may be found in Section 6.

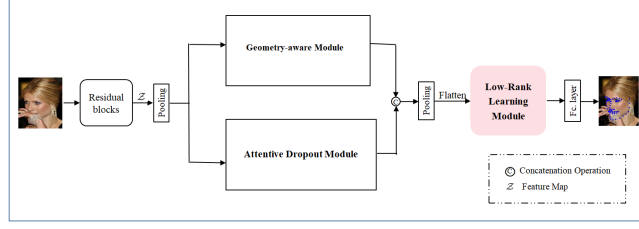


Fig. 2: Black-box diagram of proposed Attentive Dropout-based Occlusion-adaptive Deep Network (ADODN).

2 Related Background and Context

FLD aims to accurately identify predefined landmarks on facial images, facing significant challenges due to occlusion, characterized by its unpredictable and complex nature as illustrated in Figure 1. Various strategies have been developed to address occlusion, including [2] supervised regression method for updating visibility probabilities, [10] occlusion dictionary, [11] adaptive regression and shape modeling for landmark occlusion assessment, and recent advancements by [12] and [13] for precise facial point localization.

[14] initially introduced the Occlusion-adaptive Deep Network (ODN) to tackle occlusions in FLD, yet its capability was limited by insufficient feature representation and occlusion modeling. In response, [8] developed the Attentive Occlusion-adaptive Deep Network (AODN), leveraging an attention mechanism to improve feature representation. This mechanism, akin to the human visual system’s focus on specific areas for detailed observation, directs the network’s focus through weighted feature summation, addressing ODN’s limitations and enhancing the model’s ability to manage occlusions effectively.

3 Attentive Dropout-based Occlusion-adaptive Deep Network (ADODN)

In pursuit of efficient and effective discriminative feature erasure, occlusion handling, feature recovery, and improved speed and accuracy with minimal resource utilization, we propose a novel network structure. Our newly proposed model, as depicted in Figure 2 and Figure 3, comprises three closely integrated modules:

1) Geometry-aware module, 2) Attentive dropout module, and 3) Low-rank learning module.

The primary objective of this model is to distribute workloads, optimize resource allocation, and enhance both efficiency and accuracy. Initially, the feature mappings denoted as 'Z' from previous residual learning blocks are fed into the geometry-aware module and the attention-based dropout module. This facilitates the extraction of geometric information and refines the feature representation, respectively. Subsequently, the outputs from these two modules are

combined and serve as input to the low-rank learning module. This module is capable of restoring missing features by analyzing relationships among various facial components.

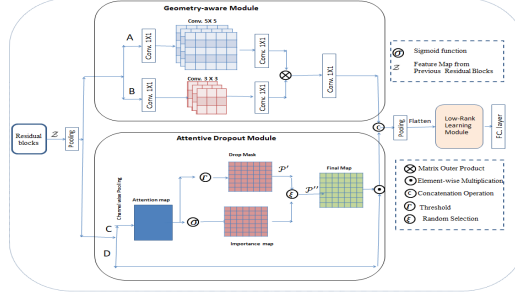


Fig. 3: The structural diagram of the proposed Attentive Dropout-based Occlusion-adaptive Deep Network.

Attention is essential for filtering occluded facial characteristics. The absence of specific features should not be misconstrued as their nonexistence. It is vital to avoid potential bias in the model’s explanation. Most facial features exhibit co-occurrence or correlation, which can aid in recovering missing attributes by considering their positional relations, proximity, or symmetry. The presence of certain characteristics can guide the retrieval of absent attributes or lead to the discovery of additional ones.

The proposed ADODN network comprises three main modules and four sub-networks. The geometry-aware module incorporates two sub-networks, labeled A and B, while the attentive dropout module encompasses sub-networks C and D. Sub-network C employs a residual block to implement Channel Attention (CA), dropout mask, and importance masking, aiming to accurately simulate occlusion and provide comprehensive, precisely localized facial feature representation. Sub-network D focuses on preserving input signal integrity for stable features. The results from C and D are combined using element-wise multiplication, assigning lower weights to background and occluded regions. Ultimately, these sub-networks generate a weighted feature map (clean features) for the entire face.

To enhance robustness and promote sparsity during optimization, we apply L_1 regularization to \mathcal{P}' and \mathcal{P}'' . The outcomes from both the geometry-aware and attentive dropout modules are combined to form a single high-dimensional feature map representing the face graph. These hybrid feature maps are then subjected to down-sampling and flattening before serving as input for the low-rank learning module. The training set $\{(I_i, \check{S}_i)\}$ can be obtained using the process described in (1).

$$\min \frac{1}{N} \sum_{i=1}^N \left\| \check{S}_i - S \right\|_F^2 + \beta \text{Rank}(\mathcal{M}) \quad (1)$$

Here, \check{S} and S represent the ground-truth and corresponding predictions, respectively. $\check{S} = \{s_1, s_2, \dots, s_L\}$, and $S = W_{fc}^T \mathcal{M}^T \mathcal{X}$. The outputs of the geometry-aware module are denoted as \mathcal{X} , where L and s denote the numbers of landmarks and facial landmarks, respectively. We employ β as a regularization factor to adjust the rank of \mathcal{M} , where W_{fc} represents the parameters of a fully connected layer. The training process for ADODN is conducted in an end-to-end manner, similar to ODN and AODN.

3.1 Attentive Dropout Module

Recent studies [15,16] have demonstrated the effectiveness of dropout layers in object detection tasks with highly efficient layers. We have also explored the utility of attention mechanisms in prior research [8,17]. Attention mechanisms typically prioritize distinctive features to enhance classification accuracy but often lead to a spatial arrangement that focuses only on the most distinctive section, resulting in decreased localization accuracy.

To enhance localization accuracy in facial landmark detection while utilizing attention mechanisms efficiently, we introduce an innovative approach featuring an attention-based dropout module. This module is designed to selectively ignore the most distinctive features in the input, thus focusing the learning process on the broader range of facial characteristics, which contributes to maintaining high classification accuracy.

This process begins with the generation of an attention map through channel-wise average pooling of the input feature map, leading to the creation of two pivotal elements: a drop mask and an importance map. The drop mask, obtained by applying a threshold to the attention map, temporarily hides the most prominent features during the training phase. This strategy prompts the model to compensate by paying more attention to the subtler, often overlooked details. Concurrently, the importance map, produced by applying a sigmoid activation to the attention map, identifies and accentuates the most informative regions, thereby bolstering the model's ability to classify facial landmarks accurately.

In practice, the model alternates between these two strategies by randomly choosing either the drop mask or the importance map in each training iteration and applying it to the feature map through spatial multiplication. This dual approach ensures a balanced learning process, emphasizing both the elimination of dominant features to diversify learning and the enhancement of critical areas to improve recognition and localization precision.

3.2 The intrinsic interconnectivity between the three modules

Our proposed framework exhibits a strong synergy among three essential modules: the geometry-aware module, attentive dropout, and low-rank learning mod-

ule. As previously noted [8,14], human visual processing encompasses two distinct pathways: the ventral stream for object identification and categorization, and the dorsal stream for processing object spatial positions. Analogously, our ADODN addresses two core aspects: occlusion awareness and geometric relationships, akin to this biological mechanism.

The geometry-aware module in our study effectively capitalizes on consistent geometric relationships such as symmetry, proximity, and location across different facial components to enhance detection capabilities. In parallel, the proposed attentive dropout module, by simulating aspects of the human visual system’s focus in network engineering, selectively screens out occluded areas and irrelevant background noise. Central to this process is the role of attention in achieving detailed feature representation, directing the network’s focus towards specific facial regions. Channel-wise Attention (CA) cues the network to prioritize crucial features of the facial image, while the incorporation of dropout masks and importance maps further sharpens localization accuracy.

The synergistic interaction between the attentive dropout and low-rank learning modules significantly improves the facial recognition feature learning process. However, despite the performance gains from these hybrid features, the selective filtering by the attentive dropout module may result in a less comprehensive representation of the entire face, especially in occluded areas. The integrated functionality of these modules enhances the overall efficacy of our Attentive Dropout-based Occlusion-adaptive Deep Network (ADODN) in overcoming obstacles presented by occlusion.

4 Optimization of Proposed Methodology

Mathematically ADODN can be formulated as minimization problem as:

$$\min \frac{1}{N} \sum_{i=1}^N \left\| \check{S}_i - S_i \right\|_F^2 + \beta \text{Rank}(\mathcal{M}) + \gamma \|\mathcal{M}\|_F^2 + \alpha \|\mathcal{W}_c\|_F^2 + \lambda \|\mathcal{W}_{fc}\|_F^2 + \eta' \left\| \mathcal{P}'_i \right\|_F^1 + \eta'' \left\| \mathcal{P}''_i \right\|_F^1, \quad (2)$$

The $\mathcal{S} = \mathcal{F}_{ADODN}(\mathcal{I}; \mathcal{W}_{fc}; \mathcal{M})$ represents the relationship between \mathcal{S} and the function $\mathcal{F}_{ADODN}(\cdot)$, which is our suggested ADODN. The symbol \mathcal{W}_c denotes the set of parameters for the convolutional layer, whereas \mathcal{W}_{fc} indicates the parameters for the fully connected layer. The parameter set of the LM is denoted by \mathcal{M} . The Frobenius norms govern the reduction in size of the three parameter sets, each associated with the connected parameters $(\alpha, \gamma, \lambda)$, respectively. The attention module applies the parameter η to impose the single-channel feature maps \mathcal{P}' and \mathcal{P}'' , which are then subjected to the L_1 operation.

In the equation where $\mathcal{S} = \mathcal{F}_{ADODN}(\mathcal{I}; \mathcal{W}_{fc}; \mathcal{M})$, and $\mathcal{F}_{ADODN}(\cdot)$ represents our proposed ADODN, \mathcal{W}_c denotes the parameter set of the convolutional layer, and \mathcal{W}_{fc} typically represents the parameters of the fully connected

layer. The inclusion of the nuclear norm $\|\mathcal{M}\|_*$ serves the purpose of addressing issues associated with low-rank learning, as it provides the most stringent lower bound among all convex lower bounds of the rank function. Therefore, the above equation can be reformulated as follows:

$$\min \frac{1}{N} \sum_{i=1}^N \left\| \check{S}_i - S_i \right\|_F^2 + \beta \|\mathcal{M}\|_* + \gamma \|\mathcal{M}\|_F^2 + \alpha \|\mathcal{W}_c\|_F^2 + \lambda \|\mathcal{W}_{fc}\|_F^2 + \eta' \left\| \mathcal{P}'_i \right\|_F^1 + \eta'' \left\| \mathcal{P}''_i \right\|_F^1, \quad (3)$$

Utilizing the property of circularity of the trace and employing the definition of the nuclear norm as per [8, 14, 17, 18] We can derive the gradient of the rank function.

$$\frac{\partial \|\mathcal{P}\|_1^F}{P_k} = \begin{cases} +1 & P_k > 0 \\ -1 & P_k < 0 \\ [+1, -1], & P_k = 0 \end{cases} \quad (4)$$

Here, P_k represents the k -th element in the set \mathcal{P} , where \mathcal{P} can take on values of \mathcal{P}' or \mathcal{P}'' . Consequently, we can affirm that this forms a directed acyclic graph, allowing for the optimization of regression loss gradients, such as the L_2 loss, through back-propagation. Moreover, all parameters can be optimized in an end-to-end manner.

5 Experimental Details

This section comprehensively evaluates the proposed framework on diverse benchmark datasets for FLD, considering various task settings. Subsection 5.1 provides initial insights into the benchmark datasets and experimental parameters used. Subsection 5.2 discusses assessment measures and training specifics for ADODN. Furthermore, the impact of system characteristics and individual component contributions to FLD performance is examined. In Subsection 5.4, an ablation study validates the effectiveness of our approach. Data augmentation involved resizing, cropping, flipping, scaling, rotating, and translating images in the training set, with standard dimensions of 224×224 . Pre-training on the ImageNet dataset, as in ODN, was also conducted for all models [19].

5.1 Datasets and Specifications

We assess the proposed FLD system’s effectiveness by evaluating its performance on multiple benchmark datasets, including 300W [20], and COFW [9], all available for research purposes. A comparative analysis is conducted against several contemporary methodologies: Sadiq et al. [8, 17], Zhu et al. [14]. Initially, our model is trained on the 300W training set and then tested on various datasets.

Table 1: The comparison results on Common-set and Full-set of 300W on basis of NRMSE ($\times 10^{-2}$).

Method	Year	Common-set	Full-set
ODN[14]	2019	3.56	4.17
ADN[17]	2019	3.52	4.14
LGSA[21]	2020	3.36	4.06
3FabRec[22]	2020	3.36	3.82
AODN[8]	2022	3.27	3.76
ADODN	2024	3.10	3.60

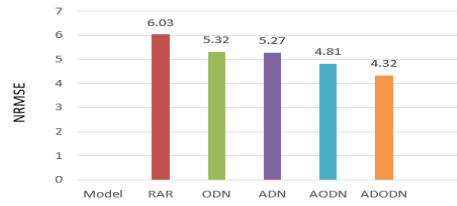


Fig. 4: Comparison of Normalized Root Mean Square Error (NRMSE) ($\times 10^{-2}$) on the COFW dataset.

- The 300W dataset, extensively used for facial landmark detection (FLD) assessment, contains 3,837 photos from AFW, LEN, and LFPW datasets, each annotated with 68 landmarks. Our training utilized 3,148 photos, with the remaining 689 for testing. The testing set is divided into: (a) Common set, with 554 photos from HELEN and LFPW datasets. (b) Challenging set, containing 135 photos from IBUG. (c) Full set, encompassing all 689 testing photos.
- The COFW dataset, publicly available, comprises 1,852 images, with 1,345 for training and 507 for testing. For our evaluation, we exclusively used the testing subset. We increased the landmark count from 29 to 68 using re-annotation following Ghiasi et al. (2014). COFW is well-known for its difficulty, featuring variations in occlusion, emotions, stance, and shape.

5.2 Metrics for Evaluation and Technical Implementation Details

We evaluated ADODN using two methods: the Cumulative Error Distribution (CED) curve and the Normalized Root Mean Squared Error (NRMSE). The NRMSE is calculated as:

Table 2: Comparison of Normalized Root Mean Square Error (NRMSE) ($\times 10^{-2}$) on the Challenging subset of the 300W dataset.

Method	Year	Challenging-set
ODN[14]	2019	6.67
ADN[17]	2019	6.60
RetinaFace[13]	2020	6.83
AODN[8]	2022	6.38
ADODN	2024	5.81

$$NRMSE = \frac{1}{N} \sum_{i=1}^N \frac{\|\check{S}_i - S_i\|_2}{L\Omega_i} \quad (5)$$

Here, L is the number of landmarks, and Ω is the inter-ocular distance, which we set to the width of the bounding box in the AFLW dataset. We experimented with different parameter settings, including reduction ratio (r) values of 8, 16, 32, and 64. After thorough analysis, we determined that $r = 8$ yielded the best results. All other parameters were kept consistent with those used in ODN and AODN.

5.3 Empirical analysis

To evaluate our method’s performance under standard conditions, we conducted assessments using two benchmark subsets from the 300W dataset: the common set and the Fullset. These subsets have minimal variations in illumination, pose, and occlusion. Table 1. shows NRMSE (normalized root mean square error) results ($\times 10^{-2}$), with comparisons against state-of-the-art models. ODN achieved an NRMSE of 3.56 for the common set, while AODN achieved 3.27. Notably, ADODN improved performance significantly, reducing the error upto 3.10 on the common set.

We also evaluated our model on the 300W Fullset, and substantial improvements can be observed in Table 1. Additionally, Figure 5(b) illustrates the Cumulative Error Distribution (CED) curve for the Fullset, demonstrating significant enhancements in results compared to other existing methods.

Table 2 presents the comparative results for the challenging set, where we have evaluated our approach against current state-of-the-art methods. A noticeable improvement is evident in the comparison table, with ADODN yielding a substantial reduction in error, from 6.67 to 5.81 ($\times 10^{-2}$) in comparison of ODN, marking a significant advancement in model performance.

Moreover, in Figure 4, we illustrate the results of ADODN for the COFW dataset, which also demonstrate remarkable improvements. Figure 5(a) depicts the Cumulative Error Distribution (CED) curve for the challenging set, showcasing significant enhancements in both CED curves.

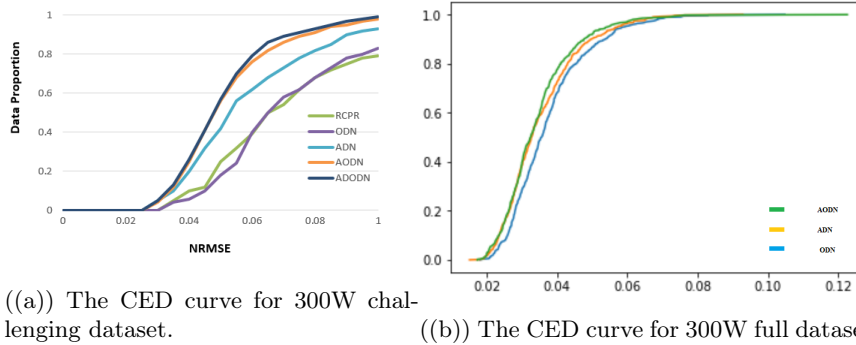


Fig. 5: The Cumulative Error Distribution (CED) Curve for the 300W Dataset.

Table 3: Ablation study conducted on the 300W Challenging dataset.

Model	NRMSE
BRNet	7.21
BRNet+GM+LM	6.90
BRNet+GM+ADM	6.72
BRNet+ADM+LM	6.68
BRNet+GM+LM+ADM(r=8)	5.81

5.4 Ablation Analysis

ADODN integrates three modules: the Geometry-aware Module (GM) for facial geometry, the Attentive Dropout Module (ADM) enhancing feature representation through human visual system-inspired attention and dropout techniques, and the Low-rank Learning Module (LM) to recover missing features. An ablation study assesses each module’s impact on difficult datasets, with systematic evaluation against the baseline ResNet-18 and modifications within ADODN, highlighting channel-wise attention’s role in feature emphasis. Table 3 quantifies the significance of each module and highlights our model’s robustness on the challenging 300W dataset, and confirming that our proposed model represents the most effective combination of these modules.

Furthermore, Figure 6 displays the qualitative detection results of our approach on select sample faces from the 300W full dataset, with ground-truth landmarks in blue (top row) and our predicted landmarks in yellow (bottom row) along with approximate error.

6 Conclusion and future Work

This paper introduces an attentive dropout-based occlusion-adaptive deep network to address Facial Landmark Detection problems. Specifically, we enhance

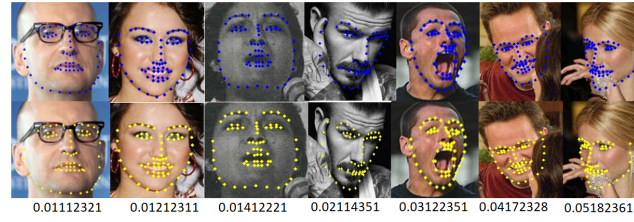


Fig. 6: Qualitative analysis of the proposed method’s detection outcomes is presented using chosen examples from the full 300W dataset. The actual landmarks are marked in blue (top row), while the landmarks predicted by our method are shown in yellow (bottom row). Additionally, the estimated error is also given

existing well established ODN, and AODN model by incorporating channel-wise attention, dropout mask, and importance map to improve its performance. Channel attention helps prioritize informative features in input facial images, while dropout mask, and importance map directs the network’s focus and improve localization accuracy. We extensively evaluate our framework on diverse benchmark datasets, comparing it to state-of-the-art methods. Results demonstrate superior accuracy in facial landmark detection. Future work includes implementing facial expression recognition using our robust model and developing a parallel computing version for enhanced efficiency. Furthermore, we will extend our work for video analysis as well.

Acknowledgments. This work is supported by the Science and Technology Ph.D. Research Startup Project under Grant No.SZIT2023KJ016, Shenzhen Science and Technology Program (Grand No. RCBS20221008093252092 and No. 20220820003203001) and Guangdong Basic and Applied Basic Research Foundation (Grand No. 2023A1515110070).

References

1. I. Kemelmacher-Shlizerman and R. Basri, “3d face reconstruction from a single image using a single reference face shape,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 2, pp. 394–405, 2010.
2. Y. Wu and Q. Ji, “Facial landmark detection: A literature survey,” *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, 2019.
3. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models-their training and application,” *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.
4. X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 2879–2886.
5. G. Tzimiropoulos, J. Alabort-i Medina, S. Zafeiriou, and M. Pantic, “Generic active appearance models revisited,” in *Asian Conference on Computer Vision*. Springer, 2012, pp. 650–663.

6. A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3444–3451.
7. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*. Springer, 2014, pp. 818–833.
8. M. Sadiq and D. Shi, "Attentive occlusion-adaptive deep network for facial landmark detection," *Pattern Recognition*, vol. 125, p. 108510, 2022.
9. X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1513–1520.
10. J. Xing, Z. Niu, J. Huang, W. Hu, X. Zhou, and S. Yan, "Towards robust and accurate multi-view and partially-occluded face alignment," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 987–1001, 2017.
11. Q. Liu, J. Deng, J. Yang, G. Liu, and D. Tao, "Adaptive cascade regression model for robust face alignment," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 797–807, 2016.
12. M. Sadiq, D. Shi, and J. Liang, "A robust occlusion-adaptive attention-based deep network for facial landmark detection," *Applied Intelligence*, vol. 52, no. 8, pp. 9320–9333, 2022.
13. J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5203–5212.
14. M. Zhu, D. Shi, M. Zheng, and M. Sadiq, "Robust facial landmark detection via occlusion-adaptive deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3486–3496.
15. J. Choe and H. Shim, "Attention-based dropout layer for weakly supervised object localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2219–2228.
16. J. Choe, S. Lee, and H. Shim, "Attention-based dropout layer for weakly supervised single object localization and semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 12, pp. 4256–4271, 2020.
17. M. Sadiq, D. Shi, M. Guo, and X. Cheng, "Facial landmark detection via attention-adaptive deep network," *IEEE Access*, vol. 7, pp. 181 041–181 050, 2019.
18. F. Nie, H. Huang, and C. Ding, "Low-rank matrix recovery via efficient Schatten p-norm minimization," in *Twenty-sixth AAAI conference on artificial intelligence*, 2012.
19. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
20. C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 397–403.
21. P. Gao, K. Lu, J. Xue, L. Shao, and J. Lyu, "A coarse-to-fine facial landmark detection method based on self-attention mechanism," *IEEE Transactions on Multimedia*, 2020.
22. B. Browatzki and C. Wallraven, "3fabrec: Fast few-shot face alignment by reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6110–6120.