RESEARCH PAPER

Facial Expression Recognition In Low Light And Occluded Environment Using Machine Learning Approaches

[1]*Rajesh Kumar Jha , [2]Dr. Brajesh Kumar
[1*]Research Scholar School of Computer Science and Engineering Sandip University, Sijoul, Madhubani, Bihar
Email Id- jha.rkjha.patna@gmail.com
[2]Associate Professor School of Computer Science and Engineering Sandip University, Sijoul, Madhubani, Bihar
Email Id- brajeshkumar@sandipuniversity.edu.in

Abstract
Facial Expression Recognition (FER) is important in the context of intelligent systems to recognize human emotions, and has been applied across human-computer interaction, healthcare monitoring, surveillance, and driver assistance systems. Although considerable improvements have been made in recent years, performance of FER systems significantly decreases in the real-life situation, especially when using low-light scenarios and face obfuscation. These influences decrease the visibility of features, distort facial information and cause unfinished or unreliable representation, which decreases recognition accuracy. This review is a narrative synthesis of machine learning and deep learning methods to FER in such a challenging environment. It discusses methods of enhancement, occlusion-sensitive models as well as state-of-the-art architectures, such as attention systems and transformer frameworks. It is seen that the joint enhancement-recognition models can be more effective in restoring the semantically meaningful features when the light is low, and attention and transformer-based models are more efficient in coping with the occlusion through focusing on informative areas of the faces. Nevertheless, one of the main drawbacks of the current studies is the absence of cohesive datasets and benchmarks that can cover both the low-light and occlusion conditions at the same time. The review adds a hierarchical taxonomy of the current methodologies and indicates a way forward in future research to create robust and practical systems of FER.

Keywords: Facial Expression Recognition (FER), Low-Light Image Enhancement, Occlusion Handling, Deep Learning, Vision Transformers
How to cite this article: Jha RK, Kumar B. Facial Expression Recognition In Low Light And Occluded Environment Using Machine Learning Approaches. Int J Drug Deliv Technol. 2026;16(58s): 1109-1120. DOI: 10.25258/ijddt.16.58s.115

1. Introduction

Facial Expression Recognition (FER) is one of the major tasks in affective computing that aims at recognizing human emotions by automatically detecting them using facial characteristics. FER systems can help machines to understand human affective states by analyzing the differences in facial muscle movements, thus enhancing more natural and intelligent interaction between humans and machines. FER has over the years developed to become more than the traditional machine learning to more sophisticated deep learning-based systems with a significant improvement in recognition accuracy under controlled conditions [1]. Such developments have made FER one of the important technologies in various fields of application.

The importance of FER is reflected in the extensive application of this concept in practice. FER has been used in human-computer interaction (HCI) where it is used to create dynamic adaptive systems responding to user emotions to improve user experience and interaction. FER finds more application in mental health, pain, and behavioral analysis in healthcare. FER is used in surveillance systems to identify suspicious or abnormal emotional signals, as well as in driver monitoring systems to identify fatigue, distraction, and emotional stress and ensure road safety. All these applications underscore the increasing relevance of FER in the contemporary smart systems [2].

Although such improvements have been made, one significant drawback of early FER systems is that they use controlled datasets, with the environmental factors like lighting, background, and visibility of faces being highly controlled. In such environments, facial features can be easily seen as well as expressions are well-defined and therefore models are very accurate. However, practical uses demand that FER systems can be deployed in unconstrained or in-the-wild environments, where the variability is far more significant. To address this shortcoming, large data sets such as AffectNet have been suggested and these can provide a range of facial images as they occur in a real-world scenario and these have variations of pose, light and occlusion [3].

Further developments in FER research have also been devoted to the importance of researching subtle and spontaneous movements of the face. Using the case of micro-expression analysis, it is revealed that the face can even convey a lot of emotional information by simply modifying the dynamics of the face, but it is highly sensitive to environmental interference. These results support the complexity of FER in the real-life scenario when both macro and micro-level expressive elements must be well-represented despite the noise and distortions [4]. Moreover, other datasets like Aff-Wild have expanded the FER study by adding continuous

*Author for Corresponden ; jha.rkjha.patna@gmail.com

emotion labels and real-life variability to assist in more realistic performance assessment of the model [5].

The next major innovation in this field is the creation of large, automated annotation systems like EmotionNet that apply machine learning algorithms to label large volumes of data on facial expressions in the wild. Such attempts have led to the presence of a variety of datasets and lots of data, yet these approaches also illustrate the nature of the problems that can arise due to variability in the real-world conditions, such as uneven lighting, occlusions, and the complexity of the background [6].

Of particular importance are the illumination variation and the problem of occlusion, among other practical FER systems issues. Low-light images deteriorate image quality and the contrast, facial detail, and noise in the image is lost and it directly affects feature extraction and classification. Equally, occlusion due to masks, glasses or hands or head poses result in partial visibility of the facial part, which causes incomplete or distorted feature representations and decreased accuracy of recognition.

These difficulties are aggravated by the presence of both low-light conditions and occlusion as is often seen in practice in night time surveillance of the scene and in-vehicle monitoring. Whereas deep learning models are highly accurate in controlled settings, they tend to be unable to generalize to such degraded inputs. Trained models on clean datasets are more likely to overfit to the ideal environments leading to a large drop in their performance when applied to the real world.

Although much work has been done on the enhancement of illumination and the ability of FER to consider occlusions, the issues are generally discussed separately. Their effect on recognition performance in combination is not understood well. Moreover, the absence of standardized metrics, which take into consideration low light and occlusion conditions limits dependable assessment and comparison of current practices.

This review will attempt to critically and systematically review machine learning and deep learning methods of FER in low-light and occluded conditions. It examines basic methodology, strength-based approaches and evaluation strategies and determines the key research gaps to assist the future researchers in developing a more reliable and efficient FER system.

2. Fundamentals of Facial Expression Recognition

The facial expression recognition (FER) systems typically have a structured pipeline, which is face detection, preprocessing, feature extraction, and classifications. All these stages are critical in defining how well a system performs, especially when there is variability in the real world and poor input conditions.

2.1 FER Pipeline

The initial step of any FER system is face detection, which involves the process of object location and isolation of the facial part of an input image or video. The early techniques relied on manually produced face-detection techniques, however, in the current day more robust techniques have been created through deep learning that can detect faces in various poses, sizes and brightness [7]. To illustrate, multitask convolutional networks can detect and align faces at the same time, and is far more precise and effective in localization compared to single-net approaches [8].

The preprocessing of the facial image is done after the image is detected. These are face alignment, resizing, illumination normalization and noise reduction operations. These measures are necessary to minimize variations that are not related to facial expression making input samples consistent. The alignment accuracy contributes to the normalization of facial landmarks, and this is essential to extract features reliably.

The second step is feature extraction, where representations of facial expressions that have a sense are obtained. In classic methods, this is via hand-designed features intended to represent texture and structural features in areas of the face. Lastly, the features are categorized into predefined emotional categories of happiness, sadness, anger, and surprise with the help of machine learning or deep learning classifiers.

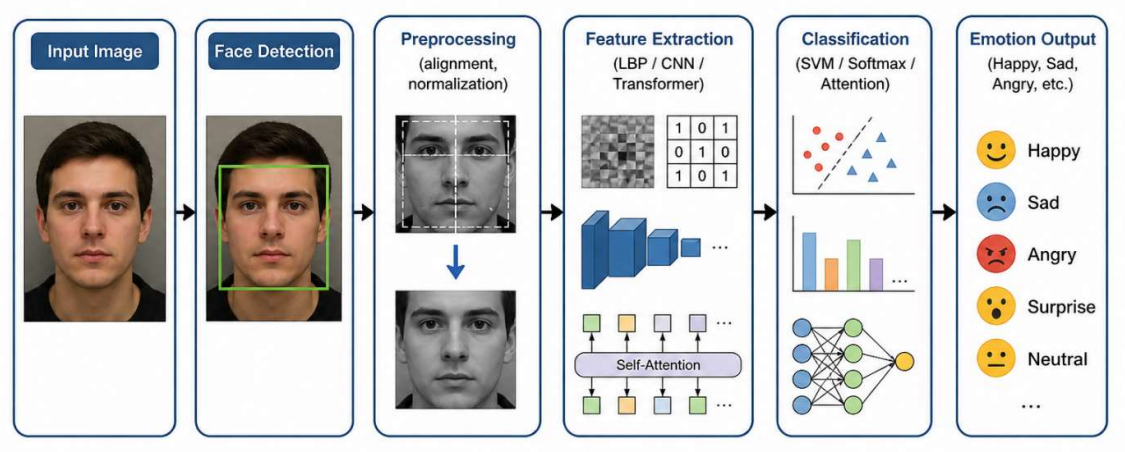

Figure 1: General pipeline of facial expression recognition (FER) systems.

As shown in Figure 1, the FER process starts with face detection, which is then followed by preprocessing which involves alignment and normalization to minimize variability. Discriminative representations of the faces are then extracted through feature extraction which can be done through either handcrafted descriptors or deep learning models. Lastly, the features of a particular emotional category are classified. This organized pipeline is the basis of most FER systems, but its performance largely relies on the strength of each step, in practice.

2.2 Traditional Machine Learning Approaches

Before the advent of deep learning, FER systems were generally founded on manual feature descriptors along with classical classifiers. One of the most used of them has been Local Binary Patterns (LBP) due to its efficiency in computation, and relative immunity to illumination variations. LBP is a local texture, which coded the differences between the intensity of the pixels, and as such, LBP is applicable in transmitting patterns of facial expressions [9]. It has been found to work on several datasets [10].

Other descriptors such as Histogram of oriented Gradients (HOG) and Gabor filters have also been used to capture edge and frequency information. These features are normally combined to improve the power to represent. It is usually classified using Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) algorithms, which are suitable when the size of data is smaller and require less computing power.

Despite its advantages, the traditional methods are not adaptable. Since features are manually drawn, they are likely to be affected by changes in light intensity, posture, and obscuration. This therefore affects their performance a great deal in free environments.

2.3 Deep Learning-Based FER

Deep learning has also made a considerable leap in FER by making learning of features out of raw data automated. It is largely controlled by CNNs such as VGG and ResNet, which can learn hierarchical representations, beginning with the low-level textures, and gradually ascending to higher-level representations of semantic features, increasing the recognition accuracy.

To capture temporal dynamics, hybrid CNNLSTM models have been developed to investigate video-based FER, and models are able to study how an expression changes with time. Architectures on transformers have subsequently suggested self-attention to capture global dependencies on face regions.

The benefits of deep learning algorithms include large data sets and improved training methods, including those that can be applied to handle noisy labels and optimize generalization [11]. However, despite the good performance, these models are susceptible to distortion of the environment such as low-light conditions and occlusion. This predisposes them to be highly reliant on large and big datasets, limiting their robustness to bad quality real-world inputs.

Table 1: Taxonomy of Facial Expression Recognition Methods

Category	Method Type	Key Techniques	Advantages	Limitations
Traditional Machine Learning	LBP, HOG, Gabor + SVM/KNN	Handcrafted feature extraction	Lightweight, interpretable	Sensitive to illumination and occlusion
Deep Learning	CNN, CNN+LSTM	Automatic hierarchical feature learning	High accuracy, scalable	Data-dependent, computationally expensive
Transformer-Based Models	Vision Transformers (ViT)	Self-attention, global context modeling	Robust to occlusion, captures global relationships	High computational cost, requires large datasets

To give a systematic picture of FER methodologies, Table 1 gives a taxonomy of key approaches. Table 1 indicates that the conventional machine learning algorithms are more focused on computational efficiency, yet not resilient to real-world complex environments. Deep learning methods are more accurate when learning features based on data but are susceptible to data quality and environmental changes. Model based on transformers, however, have a greater degree of robustness, as they consider global contextual relationships, but they have higher computational complexity. This analogy underscores the inherent efficiency, accuracy and robustness trade-offs among various FER paradigms and it drives the necessity of more robust solutions in practice.

3. Challenges in Real-World Facial Expression Recognition

Facial Expression Recognition (FER) systems have serious performance drawbacks in the realistic world. Contrary to the controlled laboratory environment, real-world conditions can vary in terms of lighting, occlusion, pose, and background conditions. Of these, the most important ones are the illumination variation and occlusion, which directly influence the visibility and integrity of facial features, interrupting the process of feature extraction and classification, and lowering the overall reliability.

3.1 Illumination Variations

The most significant issue to FER is the difference in the illumination, particularly, when it is dark. Fine-grained

*Author for Corresponden ; jha.rkjha.patna@gmail.com

facial features such as wrinkles, lines and the more subtle movements of muscles are necessary to be properly identified. However, this reduces contrast and conceals texture information, thus making it difficult to generate meaningful features by traditional and deep learning models.

The low light also introduces noise and reduces the visibility of the features simultaneously. Image sensors boost noise in dark environments, distorting pixel level data and decrease local intensity distributions. At the same time, the most important elements of the face such as eyes, eyebrows, and mouth may be covered partially or entirely. This feature loss coupled with loss of information severely impairs the representation of features and ability of models to detect discriminative expression cues.

It is proven experimentally that the dark condition of FER accuracy is low due to the low image quality and inability to see features. Research on low-light image enhancement indicates that, to obtain high-level vision tasks which are dependable, including FER it might be necessary to improve the light conditions [12]. Without the correct kind of preprocessing, models are prone to misclassify expressions or fail to pick up subtle emotional expressions, and models do not just pick up the difference in illumination but the entire FER pipeline.

In reality, low-light-conditions are usually accompanied with occlusion, which solely exacerbates the worsening of features and reduces recognition reliability. It is at this juncture that the need of FER models with the capability of taking into account a number of environmental distortions simultaneously arises.

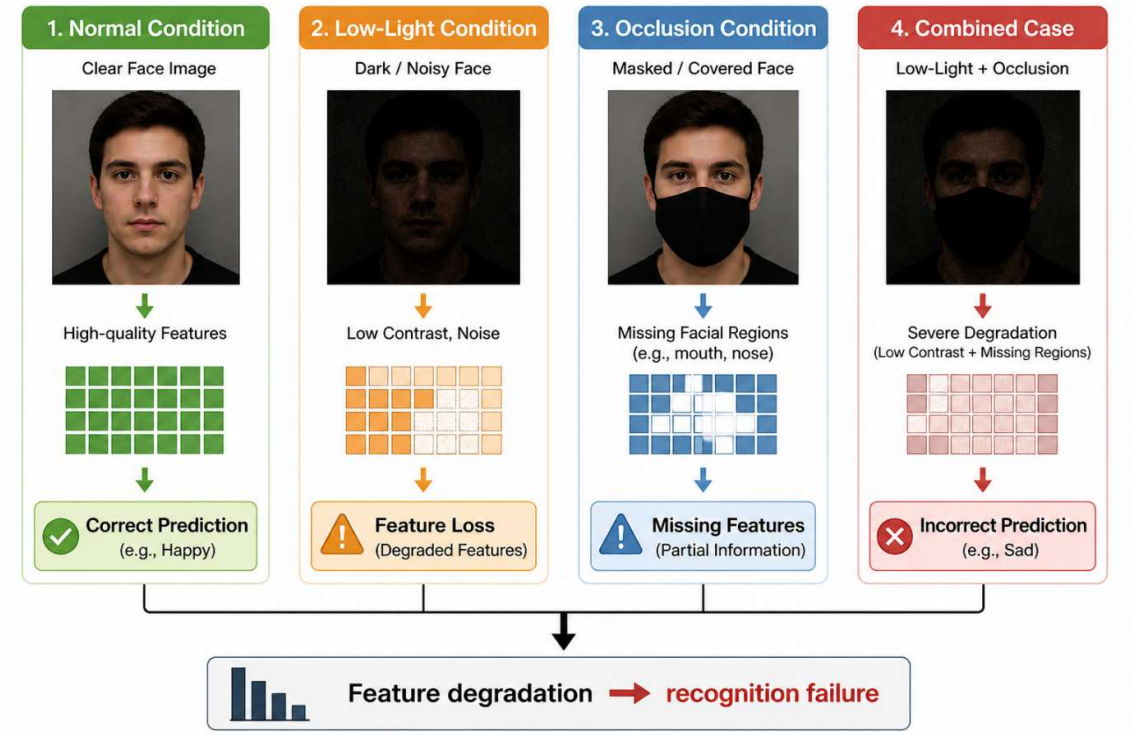

Figure 2: Impact of low-light and occlusion on facial expression recognition performance.

In Figure 2, under normal conditions, the features can be easily extracted and predictable, but in low-light environments, there is noise and a decrease in contrast, leading to a loss of features. Equally, occlusion results in the loss of facial areas, which restrict access to important expression information. Combining the two conditions together gives rise to a much worse quality of features, leading to the wrong classification, in most cases. This finding supports the necessity of the integrated FER models that will be able to solve the illumination and the occlusion problems simultaneously.

3.2 Occlusion in Facial Expression Recognition

Occlusion is one of the major issues of FER as it leads to partial or distorted faces. It occurs when the facial regions are obscured by the natural objects such as hair, hands or glasses or those which are artificial such as masks, scarves, helmets among other virtual reality instruments. Face masks are an everyday accessory, which only makes occlusion-aware FER systems that more important.

The major effect of occlusion is the loss of expression-relevant facial information. Since emotional expressions often involve some components of the face, losing facial regions can significantly reduce the recognition accuracy and force models to rely on less informative features. The traditional convolutional networks are also prone to attack because the networks normally handle the face in a holistic manner and all the areas are

presumed to be effective. This is not true with occlusion and can result in biased or incomplete representations.

To address this obstacle, region-based and attention-based strategies are proposed. The attention networks of regions dynamically attend informative and visible areas of the face [13], and the models that are conscious of the occlusion mitigate the effects of corrupted areas when extracting features [14]. However, when large areas of faces are covered it is difficult to notice it. Moreover, occlusion disrupts the recognition of the facial action units, which are important in facial expression [15]. Thus, the low-level feature extraction and high-level classification are affected by occlusion.

3.3 Combined Degradation: Low-Light and Occlusion

Whereas illumination variation and occlusion are commonly investigated independently, in practice, real-world FER systems can experience either or both. This may include low lighting conditions at night with surveillance, surveillance of drivers as well as surveillance of inside an establishment with masks, sunglasses, steering wheel or other partially concealing objects.

Their combination is worse than either one of the conditions. Low light causes loss of contrast and noise whereas occlusion eliminates vital parts of the face. This leads to the loss of quality as well as completeness of facial information, rendering the traditional preprocessing and recognition models inadequate.

One of the research gaps is the absence of datasets and benchmarks that combine low-light and occlusion conditions. Most of the current datasets would deal with these issues separately and restrict the validity of the assessment to realistic degradation. Likewise, current methods tend to address the enhancement of illumination and occlusion management as two distinct functions. Improvement techniques can make things visible but not where missing regions, whereas models that are aware of occlusions can break down in the case of missing regions that are degenerated by bad illumination.

Thus, a powerful FER needs a combined method that concurrently tackles degradation and occlusion of illumination. These involve coherent datasets, condition-conditioned evaluation protocols and models can retain, choose or recreate expression-relevant features during joint degradation.

4. Facial Expression Recognition in Low-Light Environments

Diffuse-light environments are a major challenge to FER systems because of poor visual information that is necessary to extract accurate features. Poor lighting evokes a lack of contrast, noise, and diminishes fine facial features that help to differentiate emotions. To solve this, there are two broad areas of research: improvement-based preprocessing and collaborative learning.

4.1 Image Enhancement-Based Approaches

The typical approach is to preprocess the images with the enhancement methods to restore the lost details and enhance the visibility. Simple techniques, like histogram equalization, boost contrast and can increase noise and lead to over-enhanced results which reduces their usefulness.

Greater techniques are founded on Retinex theory to split illumination and reflectance to correct lighting without destroying structure. An example of such methods is the LIME technique which approximates an illumination map to improve low-light images without changing the structure [16].

Enhancement has also been enhanced through Generative Adversarial Networks (GANs) which have been conditioned to convert low-light images to their high-light counterparts without any paired data. EnlightenGAN generates realistically looking images with better contrast, and it is useful in downstream FER tasks [17]. Retinex extensions that use deep learning, like unrolling extensions, integrates both model-based priors and data-driven learning to obtain adaptive improvement in different conditions [18]. Also, there is the integration of perceptual and structural constraints that are critical to maintain the enhancement images to be meaningful in recognition tasks [19].

Although these advances have been made, the independent methods based on enhancement are still limited, since they do not focus on the preservation of expression-specific features, but on the visual quality.

4.2 Joint Learning Models

To overcome this shortcoming, collaborative learning models combine improvement and reward into one model. These techniques enhance image restoration in an objective-driven classification context, rather than solely based on visual appearance, by incorporating enhancement modules into the FER image restoration pipeline.

These models enhance semantic feature recovery by making sure that the enriched representations are discriminative to the expression classification. They also minimize the mismatch between preprocessing and recognition steps as well as enhancing efficiency, by sharing representations across tasks. Consequently, joint enhancement recognition models are a better solution in low-light FER.

4.3 Limitations of Enhancement-Based Approaches

Although things have improved, there are still several challenges. Enhancement techniques can also create artifacts including increasing noise, over smoothing or distorting the structure, which has a detrimental effect on recognition performance. In addition, most enhancement methods do not have semantic awareness, which enhances brightness and corrupts subtle expression features.

The second limitation is that their generalization is limited by the fact that most models are trained on general image data and not on FER-specific data. Even though joint learning partially can solve this problem, it

adds extra computational complexity and reliance on large, annotated datasets.

Table 2: Comparative Analysis of Low-Light FER Methods

Method	Technique	Strength	Weakness	FER Impact
LIME	Retinex-based	Preserves structural details	Limited semantic awareness	Moderate
EnlightenGAN	GAN-based	No paired data required, visually realistic	May introduce artifacts	Good
Retinex-DL	Hybrid (model + DL)	Adaptive to varying illumination	Computationally complex	Strong
Joint Models	End-to-end learning	Semantic-aware feature optimization	Data-intensive, complex training	Best

Table 2 explains that the trade-offs of different low-light FER strategies are not similar. Algorithms such as LIME with Retinex provide stable structural preservation, but not semantics optimization. GAN-based approaches are more realistic to the eye, yet may introduce artifacts, which may undermine recognition. Hybrid Retinex-deep learning models offer enhanced degree of adaptability but at the cost of complexity. Among all the approaches, joint enhancement recognition models work best as they explicitly optimize features to facilitate expression classification, but not visual quality.

This comparison demonstrates that even though the separate enhancement procedures render the improvements visible, the combined and semantically aware models are better suited in strengthening the robust FER during low-light situations.

5. FER under Occlusion

The problem of occlusion is one of the largest in facial expression recognition since it destroys or distorts facial areas needed to discern emotions. Contrary to the illumination variation that principally reduces image quality, occlusion may completely remove expression-specific cues. The presence of masked, gloved, haired, helmeted, or glassed-in regions results in the FER systems having an incomplete face image, and thus a false classification is produced.

Illumination degradation is also well associated with occlusion in real life FER systems. A powerful model should not lose any meaningful facial features but must determine which areas of face are still valuable in recognition. It has been demonstrated that joint low-light FER systems can display images that are brighter, but this does not necessarily preserve expression-specific information [20]. This applies to the occlusion-aware FER since models should pay attention to semantically significant areas instead of considering every facial area as equal.

5.1 Region-Based Learning

The region-based learning approach deals with occlusion by splitting the face into meaningful sub-regions and concentrating on visible regions. These techniques break down the entire face as a unit representation to analyze local parts of the face, including the eyes, eyebrows, cheeks, forehead, nose and mouth. This minimizes reliance on a particular part of the face and enables the model to suppress missing/untrustworthy parts.

But region-based approaches require precise face alignment and landmark localization. Facial landmarks can be poorly detected with severe occlusion, pose variation or low-light conditions, and so region partitioning is not reliable. In addition, facial expressions can be highly co-ordinated movements in many parts of the body and therefore, local analysis may not be effective to identify them strongly.

5.2 Attention Mechanisms

Attention mechanisms are used to augment occlusion-conscious FER by assigning different weights of significance to the parts of the face. Attention-based models are trained to focus on more informative areas to classify rather than select visible parts by hand. Such adaptive weighting may be helpful to reduce the impact of part corruption as well as prioritize parts visible and expressively important.

The recognition reliability considering the compensating missing information shows that the dynamic FER algorithms under partial occlusion show higher reliability. It can be used with video-based FER, where motion features remain, because the optical-flow reconstruction, e.g., can be used to recover temporal expression data in case of partial blocking of facial areas [21]. In addition, the global and local attention strategies combine coarse facial structure and fine-grained regional features in a way that the models retain the general trends of the expression and the localized discriminative features [22].

5.3 Reconstruction-Based Methods

Reconstruction methods are those methods that attempt to re-construct or approximate lost information about the face due to occlusion. They can be generative models, latent representation learning, optical flow, or feature-level completion to create a more complete facial representation to be employed in classification. Video-based FER Temporal context can also be valuable in the estimation of the missing data in the current and future frame.

Reconstruction-based FER, however, should be cautiously employed. A visual representation of an estimated face may not be emotionally accurate. The model may give false emotional signals when it

generates incorrect patterns of mouth, eyebrows or eyebrows. This means that the process of reconstruction cannot be evaluated based on visual quality; the only thing that is of value is whether the characteristics that are recovered contribute to better expression recognition.

5.4 Transformer-Based Models

The outlook of occlusion-insensitive FER with transformer-based models is promising since self-attention can identify the relationship between remote areas of the face. Transformers, in contrast to convolutional models, encode global dependencies, and can encode non-corrupted pixels when other pixels are missing or corrupted.

Attention based loss functions are even more practical in unconstrained environment to discriminate features. The attentive center loss along with deep attention increase compactness of intra-class and separation of inter-classes, so that FER models can learn more robust representations in the face of real-world variability [23].

Nevertheless, massive data and effective computation are also necessary, based on transformer and attention models which also need to be properly designed. They are also able to streamline dataset-specific occlusion patterns or untrustworthy attention weights, in low-data circumstances.

Normally, FER under occlusions needs techniques that can detect visible areas, suppress corrupted data, interpolate missing data and maintain the overall face structure. This goal can be met through regional based, attention, reconstruction and transformer inspired but none of the strategies can be implemented comprehensively to eliminate the issue. More natural occluded cases require increasingly hybrid models comprising of region-conscious extraction, adaptive attention and semantic reconstruction models.

To present a systematic comparison of the key strategies employed in the management of occlusion in FER, Table 3 presents the key methodological categories and their trade-offs.

Table 3: Comparative Analysis of Occlusion Handling Methods in FER

Method Type	Strategy	Strength	Limitation
Region-Based Learning	Local region partitioning	Simple, interpretable, reduces dependency on full face	Requires accurate landmark detection; limited global context
Attention Mechanisms	Adaptive weighting of facial regions	Focuses on informative areas; robust to partial occlusion	Data-dependent; may mislearn attention under bias
Reconstruction-Based Methods	GAN/latent feature recovery	Recovers missing regions; improves feature completeness	Risk of generating incorrect or misleading features
Transformer-Based Models	Global self-attention modeling	Captures long-range dependencies; strong occlusion robustness	High computational cost; requires large datasets

As Table 3 indicates, both approaches have their unique strengths and weaknesses. The simplest are region-based approaches that rely on trustworthy landmarks, the adaptive attention methods are data-dependent, reconstruction method recovers missing information, semantic errors, and transformer-based models offer powerful global reasoning, but demand high computational costs. This analogy indicates that the hybrid solutions are best when it comes to strong FER under occlusion.

6. Machine Learning Approaches for Robust FER

Strong facial expression recognition (FER) has advanced beyond manually constructed machine learning pipelines to deep, hybrid, and attention-based models. In low-light and concealed conditions, robustness is not solely based on the accuracy of the classification, but it is also based on the capability of bearing incomplete, noisy or damaged facial data. This is why the primary issue is the development of models that maintain the discriminative expression signals in unfavourable visual conditions.

6.1 Traditional Machine Learning vs Deep Learning

The classic methods of machine learning involve hand-designed features like Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gabor features and classifiers like Support Vector Machines (SVM) or K-Nearest Neighbors (KNN). They are resource-limited, interpretable, and lightweight methods that are appropriate to use in FER systems. Nonetheless, they are limited by manual design of their features, which restricts flexibility. Handcrafted descriptors usually do not work when the light is lower or facial areas are covered as they are dependent on the visible texture and constant facial structure.

Deep learning models enhance representation learning by directly learning hierarchical features by extracting data. Nevertheless, they are not necessarily strong. They are highly reliant on training data diversity and can overfit to dataset-specific conditions when the low-light, occlusion or cases of compound degradation cases are underrepresented. Therefore, classical ML is effective and non-adaptable whereas deep learning is robust and susceptible to domain changes.

6.2 Hybrid Models

Hybrid models that mitigate the difference between accuracy/robustness combine complementary components, including CNNs, attention blocks, and temporal networks. Attention CNN-attention models Attention models can be particularly advantageous in

occluded FER because attention ensures that the network focuses on the informative parts and represses corrupted parts. Dual-attention cross-fusion networks channels and spatial attention are the attention that amplifies the representation of the model to highlight the characteristics of expression-relevant in the noise and occlusion presence [24].

The other hybrid is the combination of CNNs with the time models like Long Short-Term Memory (LSTM) networks. FER is video based and makes use of expressions whose formation varies with time and CNNLSTM models take advantage of temporal continuity to predict the weakened or missing cues by the surrounding pictures. It can be applied during case of transitory occlusion or when the light is different between frames.

Multi-head attention is as well implemented to enhance hybrid FER by training to correlate across faces. Multi-head attention involves numerous expression-related dependencies to enhance discrimination instead of considering one area as is the case with noisy inputs or partially occluding inputs [25].

6.3 Vision Transformers

Vision Transformers add the global self-attention to FER which learns the relationships between patches of face images. This can be applied where the expression cues are still available in non-occlusively areas, despite such other areas of the face being damaged. Transformer-based FER models can attend to visible and discriminative regions selectively and retain global facial information. Attention-enhanced FER models reported recently are observed to exhibit a better level of feature selectivity and contextual representation on a variety of datasets [26].

Transformer models are however data-intensive, computationally expensive, and require close regularization. They can overfit dataset specific occlusion patterns, or illumination patterns, in low-data FER settings. Thus, transformers hold promise and need to be accompanied by sufficient data variety and effective model architecture.

6.4 Multimodal Approaches

Multimodal FER boosts robustness through the assistance of complementary data of RGB, infrared, thermal, depth or audio data. It is especially the case in low-light recognition where RGB images may lose details in terms of contrast and texture. The IR and thermal modalities are not as dependent on visible stimuli and can retain face structure during dark conditions as compared to depth data which can be used in instances of poor texture.

The advantage of multimodal systems is mainly compensation: in the situations when one modality is not reliable, a second one can provide some additional information. However, sensor cost, synchronization, high computational requirement and small size of large multimodal FER datasets are also linked to multimodal FER.

Overall, successful FER requires strategies of perfection, effectiveness, and flexibility. Hybrid attention-based models, transformer-inspired systems, and multimodal systems are the most promising directions, in combination with task-aware improvement and other training data.

7. Datasets and Evaluation Metrics

Data and metrics of evaluation are essential in the design and testing of FER systems. The model reliability is not only based on architecture, but also on dataset diversity, quality and realism. In low-light and occluded settings, this is particularly critical, since many of the traditional datasets lack sufficient coverage of such demanding conditions.

7.1 Standard Datasets

Various benchmark datasets are popular in FER. Extended CohnKanade (CK+) is a dataset that contains controlled expression sequences and is not useful in evaluating the baseline but has no real-world variability. FER2013 is more diverse with in-the-wild grayscale images, but it is noisy-labeled and poor in quality. RAF-DB and AffectNet also enhance realism by providing large-scale datasets with variations in pose, lighting and background. Nevertheless, both of them still do not provide a systematic representation of extreme low-light conditions and organized scenarios of occlusions.

Table 4: Comparative Analysis of Common FER Datasets

Dataset	Type	Lighting	Occlusion	Size	Limitation
CK+	Controlled	Good	No	Small	Not representative of real-world conditions
FER2013	In-the-wild	Moderate	Limited	Medium	Noisy labels, low image quality
RAF-DB	Real-world	Varied	Some	Large	Limited occlusion diversity
AffectNet	Large-scale in-the-wild	Varied	Some	Very large	Limited focus on low-light conditions

Table 4 demonstrates that the existing datasets are of various scale and realism, and do not completely represent a combined set of low-light and occlusion conditions, which is the reason behind the discrepancies between benchmark testing and real application.

7.2 Limitations of Existing Datasets

Majority of FER datasets are biased towards bright environments and with no extreme changes in illumination. As a result, the models, which are trained on these data, fail to work in conditions, like nighttime surveillance or indoor scenarios with low lighting.

Likewise, occlusion diversity is small. Although a few datasets contain natural occlusion, they do not often cover the various types of occlusion or extents of severity in a systematic manner limiting reliable performance assessment.

7.3 Emerging Datasets and Synthetic Augmentation

To remove these limitations, recent research have used occlusion-based datasets and artificial augmentation techniques. To reproduce poor conditions, they add to the training artificial occlusion (e.g. masks, glasses, random blocks), and changes in lighting conditions. But effective weakly designed synthetic data, even though such exist, may not have in generalization what gives them the real-world.

Alternatives Multimodal learning offers alternative modalities like infrared, thermal or depth data, which enhance flexibility to unfavorable conditions [27].

7.4 Evaluation Metrics

Accuracy, precision, recall, F1-score, and confusion matrices are the common metrics that are used in the measurement of FER performance. Unlike accuracy, which gives an approximate measure, it can be inaccurate with skewed data. Precision, recall and F1-score are more useful because they can be applied to each of the classes, and the confusion matrices illustrate the trends of misclassifications.

In order to have a strong FER, the test should not be only evaluative on aggregate accuracy but also should be tested to also measure performance in specific situations such as, low-light, occlusion and combined degradation. In practice, other performance metrics, such as latency, computing cost, and real-time reliability, are also significant, particularly in safety-critical systems, including driver monitoring [28].

8. Comparative Analysis of Methods

The comparison of the situation indicates that the issue of facial expression recognition (FER) in dark and covered conditions cannot be effectively resolved within the frames of one approach. The advantage of conventional feature-based methods, enhancement pipelines, deep learning models, attention mechanisms, reconstruction plans and transformer-based architectures have their own advantages yet are constrained by their inability to work in the degraded realistic environments. The dependence of powerful FER on the effectiveness of the method in preserving, prioritizing or restoring expression-relevant information under partial visibility therefore depends on the method.

The ancient artisan techniques enable a convenient point of reference. Spatiotemporal cues such as local binary patterns record local variations in the texture and low-level motion features and are therefore effective in controlled environments [29]. Nevertheless, they are mostly dependent on the quality of images. Noises and disappearance of contrast reduce the reliability of the descriptors in the low-light situation and elimination of the necessary facial areas by occlusion restricts their application in the real world.

Enhancement-based methods aim at enhancing visibility in pre-recognition. They are also able to extract features, recovering brightness and contrast, without changing downstream classifiers. However, the tools are predominantly oriented towards visual quality, and not semantics of expression. The enhancement would then modify more delicate facial data that would be employed in classification activities, irrespective of an increased perceptual clearness [30].

End-to-end models address this weakness by learning both the features representations and classification problems. This would also limit the difference between the process of improvement and reward and thereafter they can streamline the work. However, combined low-light and occlusion situations are extremely reliant on very massive and varied training sets that may not be at hand.

Significant difference is noted between CNN-based and transformer-based models. Small receptive fields ensure CNNs are susceptible to corrupted regions because it has a good ability to acquire local spatial properties. Transformer-based architectures, on the other hand, use self-attention to learn how various portions of the faces relate with one another so that different portions of the faces can selectively attend to information-rich and visible portions of the faces. This makes it more resistant to occlusion, and computationally and data expensive.

Other design philosophies are also said to be a reconstruction and attention-based approach to design. Reconstruction techniques are used to rebuild absent regions of a face that may enhance information on features, but too much to the cost of a false or biased information. Attention-based models on the other hand are more interested in good regions and do not make reconstruction attempts thus are more resilient to the unpredictable.

Overall, low-light and occlusion are two problems that cannot be addressed in an absolute fashion. This has been facilitated by visibility improvement capabilities, at the cost of semantics, but the deep learning models must be designed with care so that they can allow semantics generalization when the system is in a degraded state. The loud performance of CNNs is an effective performance at a bottom level, and transformer performance is more effective owing to the ability of modeling world-wide contexts. What research is showing the most potential then is exploring involving hybrid-structures where the utilization of enhancing-recognition learning of jointly implementing the attention systems would be reached where superior visibility and selective use of features would be realised. A blend of these solutions would also be subsequently applicable in the practical applications of FER in real-life where different degradations may co-exist in the applications.

9. Challenges and Research Gaps

Although the notion of facial expression discrimination (FER) has made immense strides, there has always been one underlying problem which has never been tackled and that is the low-light and the conditions of larger than

head sizing. Wise is its lack of condition-sensitive datasets, complete of the behavior of degradation of illumination and occlusion. Majority of the available benchmarks evaluate these factors individually that lead to disjointed evaluation plans. This implies that the models that are to be optimized on these datasets would not be true of the real-world performance where there is a lot of degradation interacting and degrading other complexities.

Failure in domain generalization is another important aspect. Deep learning models are very reliable under controlled settings or slightly perturbation but grossly decreases when the distribution undergoes change. The variations in the training and deployment scenarios are because of the variation of the lighting, sensor characteristics, pose and the demographic. A more fundamental restriction can be seen in the following way: the existing models can not learn expressive codes of the data rather they can only learn data-specific relationships. Thus, Vigor is exaggerated in the laboratory and deemphasized in practice.

The less talked about yet related closely with the disadvantaged problem is the semantic inconsistency of recognition and improvement of problems. Better pipelines enhance visual qualities of them, such that they do not need to preserve those features that are relevant to expressions. It may result in the cases when it is possible to consider the images as being visually superior and the essential emotional content may be overruled or repressed. The lack of a good task-sensitive optimization leads to miscorrelation of the low-level restoration and high-level recognition targets and undermines the reliability of a complete system.

Computational constraints further restrict practical deployment. Transformer-based and hybrid architectures offer improved robustness but introduce significant computational overhead. This creates a trade-off between accuracy and efficiency, particularly in edge environments such as mobile devices and in-vehicle systems. Current research lacks scalable solutions that maintain robustness under degraded conditions while meeting real-time processing requirements.

Moreover, information asymmetry and population imbalance is also a persistent problem. FER datasets will frequently be biased with regards to some expressions and population categories and this will result in uneven performance across the emotion categories and user demographics. This not only has an impact on model reliability, but also casts doubt on fair use and ethical usage, particularly in sensitive uses like healthcare and surveillance.

Lastly, condition-specific assessment models are lacking. The aggregate accuracy is reported in most of the studies without quantifying performance under controlled conditions of degradation (such as to different levels of illumination or degree of occlusion). This inhibits viable comparison of techniques and hides real soundness. To solve this missing gap, a set of uniform benchmarks and assessment procedures should be developed to explicitly model actual degradation drivers in the real world.

All together, these obstacles demonstrate that the existing FER studies remain, to a great extent, optimised in the controlled settings. To move to the practical implementation, a transition to the integrated modeling, condition-conscious assessment and leveraging representation learning needs to be made which fully takes into consideration the joint environmental degradations.

10. Future Research Directions

Further research on FER in the future needs to be directed at creating more robust, efficient and adaptable methods to the real world. Self-supervised learning is one of these promising directions, as it does not depend on huge and annotated datasets. Self-supervised models can learn meaningful representations of unlabeled data, enabling better generalization and adaptation in a wide range of settings, such as low-light and occluded ones.

The other direction that holds significance is the development of multimodal FER systems. Adding various data modalities, e.g., RGB images, infrared (IR), thermal imaging, and depth data, can contribute greatly to robustness. As an example, the infrared and thermal sensors are not as sensitive to changes in illumination and are, therefore, effective in low-light settings. A combination of these modalities with conventional RGB data enables models to complementary capture information, enhancing recognition accuracy in adverse conditions.

There is also the critical area of research in optimization of FER models to be used in deployment of edges. To achieve real-time execution on resource limited devices, lightweight architectures, model compression methods, and efficient inference strategies are required. This is particularly relevant in applications that require latency and energy efficiency like driver monitoring and mobile-based emotion recognition.

The other new direction is the creation of explainable FER systems. Since FER finds more applications in sensitive domains, including healthcare and security, it is crucial to know how models make decisions. To enhance the transparency and confidence in FER systems, explainable AI methods can be used to understand which areas or features of a face are used to make predictions.

Lastly, issues of ethics must be discussed as the FER technology is increasingly extended. There are concerns about privacy, consent and the possible abuse in systems of surveillance. To have responsible use of FER, the ethical guidelines, secure data handling procedures, and anti-bias and discrimination mechanisms should be developed.

11. Conclusion

The idea of facial expression recognition has developed tremendously since the early machine learning methods using hand engineered features to elaborate deep learning features that can learn complicated features. This has developed into tremendous advances in recognition accuracy, especially in controlled conditions. Yet, in practice, deployment creates

problems of illumination variation and occlusion, which remain to degrade system performance. This review has discussed the main methods employed to solve these problems, such as image enhancement strategies, attention-based networks, reconstruction strategies, and transformer-based networks. Though the new paradigm of FER is the deep learning, it is exposed to environmental distortions. Transformer-based model and joint learning frameworks show a promising future with becoming more robust by utilizing global context and leveraging improvement and recognition. Although these developments have taken place, several challenges have not been addressed. The lack of integrated datasets, which have the ability to capture low-light and occlusion conditions, the ability to generalize to different contexts, and the ability to run in real-time is still a hindrance. Besides, the concerns of imbalance and bias in data and ethical concerns are also indicative of more responsible and inclusive FER research. The second move to this end will be the development of effective, powerful and morally sound systems. The approaches to self-supervised learning, multimodal integration, and explainable models are likely to play a vital role in bringing FER to more practical and real-life applications.

References

[1] Shan Li and Weihong Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2022.

[2] Muhammad Sajjad, Fath U Min Ullah, Mohib Ullah, Georgia Christodoulou, Faouzi Alaya Cheikh, Mohammad Hijji, Khan Muhammad, and Joel J. P. C. Rodrigues, "A Comprehensive Survey on Deep Facial Expression Recognition: Challenges, Applications, and Future Guidelines," *Alexandria Engineering Journal*, vol. 68, pp. 817–840, 2023.

[3] Ali Mollahosseini, Behzad Hasani, and Mohammad H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019.

[4] Xianye Ben, Yi Ren, Junping Zhang, Su-Jing Wang, Kidiyo Kpalma, Weixiao Meng, and Yong-Jin Liu, "Video-Based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5826–5846, 2022.

[5] Dimosthenis Kollias, Panagiotis Tzirakis, Mihalis A. Nicolaou, Athanasios Papaioannou, Guoying Zhao, Björn Schuller, Irene Kotsia, and Stefanos Zafeiriou, "Deep Affect Prediction in-the-Wild: Aff-Wild Database and Challenge, Deep Architectures, and Beyond," *International Journal of Computer Vision*, vol. 127, no. 6–7, pp. 907–929, 2019.

[6] Carlos Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Aleix M. Martinez, "Emotionet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 42–56, 2019.

[7] Stefanos Zafeiriou, Cha Zhang, and Zhengyou Zhang, "A Survey on Face Detection in the Wild: Past, Present and Future," *Computer Vision and Image Understanding*, vol. 138, pp. 1–24, 2015.

[8] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[9] Guoying Zhao and Matti Pietikäinen, "Dynamic Texture Recognition Using Local Binary Patterns With an Application to Facial Expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[10] Caifeng Shan, Shaogang Gong, and Peter W. McOwan, "Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[11] Shan Li, Weihong Deng, and JunPing Du, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019.

[12] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy, "Low-Light Image and Video Enhancement Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 9396–9416, 2022.

[13] Kai Wang, Xiaojiang Peng, Jianfei Yang, Debin Meng, and Yu Qiao, "Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4057–4069, 2020.

[14] Yong Li, Jiabei Zeng, Shiguang Shan, and Xilin Chen, "Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, 2019.

[15] Brais Martinez, Michel F. Valstar, Bihan Jiang, and Maja Pantic, "Automatic Analysis of Facial Actions: A Survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 325–347, 2019.

[16] Xiaojie Guo, Yu Li, and Haibin Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.

[17] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang, "EnlightenGAN: Deep Light Enhancement Without Paired Supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.

[18] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo, "Retinex-Inspired Unrolling With Cooperative Prior Architecture Search for Low-Light Image Enhancement," *IEEE Transactions on Image Processing*, vol. 31, pp. 1715–1728, 2022.

[19] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang, "Beyond Brightening Low-Light Images," *International Journal of Computer Vision*, vol. 129, pp. 1013–1037, 2021.

[20] Yuanlun Xie, Jie Ou, Bihan Wen, Zitong Yu, and Wenhong Tian, "A Joint Learning Method for Low-Light Facial Expression Recognition," *Complex & Intelligent Systems*, vol. 11, no. 2, article 139, 2025.

[21] Delphine Poux, Benjamin Allaert, Nacim Ihaddadene, Ioan Marius Bilasco, Chaabane Djeraba, and Mohammed Bennamoun, "Dynamic Facial Expression Recognition Under Partial Occlusion With Optical Flow Reconstruction," *IEEE Transactions on Image Processing*, vol. 31, pp. 446–457, 2022.

[22] Zengqun Zhao, Qingshan Liu, and Shanmin Wang, "Learning Deep Global Multi-Scale and Local Attention Features for Facial Expression Recognition in the Wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 6544–6556, 2021.

[23] Amir Hossein Farzaneh and Xiaojun Qi, "Facial Expression Recognition in the Wild via Deep Attentive Center Loss," *IEEE Transactions on Affective Computing*, vol. 14, no. 2, pp. 1595–1609, 2023.

[24] Fan Zhang, Gongguan Chen, Hua Wang, and Caiming Zhang, "CF-DAN: Facial-Expression Recognition Based on Cross-Fusion Dual-Attention Network," *Computational Visual Media*, vol. 10, no. 3, pp. 593–608, 2024.

[25] Zhengyao Wen, Wenzhong Lin, Tao Wang, and Ge Xu, "Distract Your Attention: Multi-Head Cross Attention Network for Facial Expression Recognition," *Biomimetics*, vol. 8, no. 2, article 199, 2023.

[26] Jiawei Mao, Rui Xu, Xuesong Yin, Yuanqi Chang, Binling Nie, Aibin Huang, and Yigang Wang, "POSTER++: A Simpler and Stronger Facial Expression Recognition Network," *Pattern Recognition*, vol. 157, article 110951, 2025.

[27] Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019.

[28] Khan Muhammad, Amin Ullah, Jaime Lloret, Javier Del Ser, and Victor Hugo C. de Albuquerque, "Deep Learning for Safe Autonomous Driving: Current Challenges and Future Directions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4316–4336, 2021.

[29] Xiaohua Huang, Shan Li, Xiaopeng Liu, Guoying Zhao, Xiaolan Fu, and Matti Pietikäinen, "Discriminative Spatiotemporal Local Binary Pattern With Revisited Integral Projection for Spontaneous Facial Micro-Expression Recognition," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 32–47, 2019.

[30] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo, "Kindling the Darkness: A Practical Low-Light Image Enhancer," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 1, pp. 1–22, 2022.