

## RESEARCH ARTICLE

# A Unified Model for Style Classification and Emotional Response Analysis

CHU-ZE YIN 

SWJTU-LEEDS Joint School, Southwest Jiaotong University, Chengdu, Sichuan 611756, China

e-mail: yinchuze@my.swjtu.edu.cn

**ABSTRACT** The emergence of Convolutional Neural Networks (CNNs) and Vision Transformers (ViT) has markedly transformed the field of image classification and analysis, especially within the realm of computer vision. This advancement has significantly impacted various sectors, including medical diagnostics and autonomous driving, while also fostering novel intersections with artistic exploration. Despite these advancements, the challenge of seamlessly integrating art style classification with emotion prediction remains. The complex interplay between an artwork's style and the emotional reactions it triggers requires a refined methodology to accurately encapsulate this dynamic relationship. Addressing this challenge, our study presents a Unified Model for Art Style and Emotion Prediction (ASE), which adopts a multi-task learning approach. This model is structured around three main elements: Artwork Style Classification, Emotion Prediction for viewers of art, and a Task-Specific Attention Module. By incorporating a pre-trained image encoder alongside a task-specific attention mechanism, our framework facilitates the concurrent processing of multiple tasks, while honing in on specialized feature representations. The efficacy of our model is validated through the Artemis dataset, demonstrating its proficiency in both precise art style classification and the identification of emotional responses. This highlights its capability to navigate the complex relationships present within artworks effectively.

**INDEX TERMS** Art analysis, multi-task learning, style classification, emotion detection, attention mechanism.


## I. INTRODUCTION

The advancement of Computer Vision, especially with the development of Convolutional Neural Networks (CNNs) [1], [2] and Vision Transformers (ViT) [3], has marked a new era in image classification. This technological leap has made these tools exceptionally proficient in extracting intricate semantic content, revolutionizing fields like image segmentation and target detection. Their applications extend beyond traditional boundaries, significantly impacting areas from medical diagnostics to autonomous driving systems, demonstrating a profound shift in how computers process visual data.

In the realm of art, these technological strides have created a unique intersection between computational capabilities and artistic exploration. Computer Vision tasks [4], [5],

[6], [7], [8] such as style classification, emotion detection, image restoration, and color analysis have become integral in understanding and interpreting art. These tools have empowered researchers and artists alike, offering new methods for categorizing artistic styles, deciphering the emotions conveyed through art, restoring aged or damaged artworks, and analyzing the intricate color dynamics present in various art forms.

The evolution in image classification and analysis not only enhances our comprehension of visual data but also introduces new prospects in the realm of multi-task learning [9], [10], [11], particularly in the art world. Here, the joint training in tasks such as style classification and sentiment prediction in artworks is a testament to the efficient utilization and transfer of knowledge across different but related tasks. This methodology promotes the development of a shared, robust, and versatile representation, augmenting the model's proficiency in analyzing art images. It considers the intricate

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval .

relationship between different aspects of art, like the interplay of style and emotion, offering a comprehensive view of artistic expressions.

Building on this concept, our research introduces a Unified Model for Art Style and Emotion Prediction (ASE). This model harnesses multi-task learning to address two key tasks simultaneously: the classification of artwork style and the classification of emotions in artwork viewers. Specifically, our model comprises three main components: Artwork Style Classification, Emotion Prediction for Artwork Viewers, and a Task-Specific Attention Module. The Artwork Style Classification component focuses on identifying the stylistic attributes of the art, while the Emotion Prediction segment is dedicated to discerning the emotional responses elicited in viewers of the artwork. The Task-Specific Attention Module is a critical element that integrates and optimizes the learning process across these tasks, ensuring that each aspect contributes to a comprehensive understanding of the artwork in question.

The core of our experimental model lies in its unique architecture. Built on a pre-trained image encoder, it incorporates a task-specific attention mechanism tailored for both style classification and sentiment analysis. This design allows our model to adeptly handle multiple tasks, learning specialized feature representations unique to each. The attention mechanism plays a crucial role, harmonizing features from both tasks and fostering adaptive learning, which enhances the model's multi-task capabilities.

To rigorously test and validate our model, we selected the ArtEmis dataset [12]—a diverse and comprehensive collection of artworks encompassing a vast range of styles and emotional expressions. This dataset provides an ideal testing ground, challenging our model to discern and understand the nuanced interrelations between artistic style and emotional response. Through this application, we aim to showcase the model's effectiveness in capturing these complex relationships. The insights derived from applying our model to this dataset have the potential to significantly influence the fields of art history, curation, and creation, offering novel perspectives on how we perceive and engage with art.

In this paper, our contributions are as below:

- In this paper, we introduce a novel model that effectively combines style classification and emotion detection in artworks. This dual-focused approach marks a significant advancement in understanding the intricate relationship between an artwork's style and the emotions it evokes.
- Methodologically, our approach is distinguished by the use of a pre-trained image encoder, enhanced with a task-specific attention mechanism. This mechanism is pivotal in handling the multi-task challenge of simultaneously classifying style and detecting sentiment in artworks.
- In the experiments, our method has been rigorously evaluated using the ArtEmis dataset, which offers a

diverse array of artistic styles and emotional expressions. Our results demonstrate the model's proficiency in accurately classifying styles and detecting emotions, providing valuable insights for art historians, curators, and creators.

This paper is structured to clearly present our research. After this introduction, Section II reviews relevant literature on Computer Vision in art analysis, contextualizing our work within current technological advancements. Section III details the architecture of our Unified Model for Art Style and Emotion Prediction (ASE), explaining its key components. Section IV describes our experimental methodology, covering the dataset, training procedures, and evaluation metrics, and presents our experimental results, highlighting the effectiveness of our model. Finally, Section V summarizes our main contributions.

## II. RELATED WORK

### A. IMAGE CLASSIFICATION

The introduction of AlexNet in 2012 catalyzed the widespread adoption of convolutional neural networks (CNNs) for image processing tasks [13]. This was followed by the development of the Residual Attention Network in 2016, which melded the depth of the VGG network with the computational efficiency of GoogLeNet [14], laying a foundational stone for future frameworks. Enhancements such as ResNeXt and ResNeSt models [1], [2] brought forward channel-wise attention and multi-path representation, markedly boosting the precision and efficiency of image classification. A further innovation was seen with models incorporating the Convolutional Block Attention Module (CBAM) [15], which significantly augmented the network's representational capacity by adaptively refining feature maps through sequential channel and spatial attention mechanisms. This has shown superior performance in varied tasks, including Remote Sensing Image Change Detection [16] and Recognition of Fly Species [17], while also enhancing model interpretability—a key aspect for sentiment prediction in our research. Non-local blocks, as introduced by [18], have broadened the application spectrum to include image restoration [19] and semantic segmentation [20], excelling at addressing long-range dependencies crucial for art image analysis. Integrating attention mechanisms with strip pooling [21] has further advanced scene analysis by effectively capturing extensive contextual information without sacrificing local detail fidelity. The Vision Transformer (ViT) continues to demonstrate competitiveness in image classification. Extensions like Transformer iN Transformer (TNT) [3], which processes images into visual sentences and words for finer detail extraction, and Query2Label [22], utilizing cross-attention for feature aggregation, prove particularly adept for multi-label art image classification.

### B. ARTISTIC IMAGE STYLE CLASSIFICATION

Art images, with their subjective nature, complexity, and rich semantic layers, present unique challenges for classification.

Traditional methods falter due to the distinct textures and color schemes of art compared to photographic imagery. CNNs have shown utility, with transfer learning approaches achieving significant accuracy improvements in art image classification, as seen with ResNet50 and Inception V1 networks pre-trained on ImageNet and applied to artworks [4], [5]. The use of CaffeNet for style, genre, artist, and nationality classification [23] and its adaptation as a similarity feature extractor highlights the adaptability of neural networks to art imagery. The transformer architecture, with its self-attention capability, is particularly suited to capturing the nuances of art images, facilitating tasks like style transfer [24], [25]. Comparative studies of CNNs and Vision Transformers (VTs) on datasets like the Rijksmuseum Challenge highlight the superior performance of VTs in art-related classifications [26], [27], emphasizing the efficiency of VTs in handling complex image data. Our research seeks to extend these methodologies by integrating multi-task learning for joint style and emotion prediction in art images, addressing the limitations of single-dimensional classifications.

### C. EMOTION PREDICTION

Emotion classification in art has traditionally been bifurcated into negative and positive categories, employing both sentiment polarity and affective model-based classifications [6], [7], [8]. The former utilizes classifiers to discern sentiment levels, applying algorithms and techniques focused on color, texture, and shape [28], [29], [30], [31]. The latter categorizes emotions into affective states, leveraging sentiment models for deep learning-based analysis [32], [33], [34], [35]. Multimodal sentiment analysis combines various data modalities, enhancing the depth of sentiment analysis [36], [37]. Datasets like IAPS facilitate this with a broad range of images and sentiment ratings [38]. Our approach utilizes binary and multi-classification strategies for comprehensive emotion analysis in art, employing metrics like accuracy, precision, and recall for binary tasks and confusion matrices for multi-classification tasks [39], [40].

### D. MULTI-TASK LEARNING

Multi-Task Learning (MTL) has revolutionized the design and training of deep neural networks by improving data utilization, reducing overfitting, and enhancing model efficiency [10], [41], [42], [43]. It has shown efficacy in diverse computer vision applications, from hyperspectral image classification to malware detection and fine-grained image analysis [44], [45], [46]. A significant challenge in MTL is balancing task specificity with shared learning. The Multi-Task Attention Network (MTAN) addresses this by facilitating the automatic learning of shared and task-specific features [47], [48]. Its architecture, adaptable to various feed-forward networks, optimizes feature selection and task weighting, proving its value in applications like tumor segmentation and benchmark datasets [49], [50].

## III. METHOD

In this section, we provide a comprehensive overview of the methodology employed for the joint classification of emotions and artistic styles in art images. We propose a Unified model for Art Style and Emotion prediction (ASE). This approach employs multi-task learning to concurrently address two pivotal tasks: Artwork Style Classification and Emotion Prediction for Artwork Viewers. Specifically, the model comprises three components: one for Artwork Style Classification, another for Emotion Prediction for Artwork Viewers, and a third for the Task-Specific Attention Module.

### A. ARTWORK STYLE CLASSIFICATION

In the Artwork Style Classification process, we first consider the input image as  $X$ . We use the visual coder  $f_{\text{encoder}}$  to perform feature extraction on the input image  $X$  to obtain the feature map  $h$ . This can be expressed as:

$$h = f_{\text{encoder}}(X) \quad (1)$$

Feature extraction is performed using pre-trained models such as ResNet, MobileNet, etc. because of their excellent performance on image classification tasks with strong feature learning capabilities. The feature map  $h$  is subsequently mapped to the category probability distribution of Artwork Style through a linear layer  $g_{\text{mlp}}$ . This can be expressed as:

$$P(S|X) = \text{softmax}(g_{\text{mlp}}(\text{flatten}(h))) \quad (2)$$

where  $S$  is the category of Artwork Style,  $P(S|X)$  is the probability that Artwork Style is  $S$  given the input image  $X$ , and  $\text{flatten}(\cdot)$  spreads the feature map into a one-dimensional vector, which is then mapped through the linear layer  $g_{\text{mlp}}$ . Finally, the category probability distribution is obtained using the softmax function. The extracted features are mapped to the category probability distribution through a linear layer. This is designed to learn the abstract representation of the input image and get the probability of each category by softmax function. The difference between the predicted probability distribution and the actual labels is calculated using the cross-entropy loss function:

$$\text{Loss}_{\text{style}} = \text{CrossEntropy}(P(E|X), GT) \quad (3)$$

where  $GT$  are the real category labels of the art style. The cross-entropy loss function is chosen because it is a commonly used loss function in classification tasks that measures the difference between the probability distribution of the model output and the actual labels.

The optimisation process uses Adam optimiser for parameter updating with adaptive learning rate and weight decay. Overall, the process is an image classification task through transfer learning and end-to-end supervised learning, which utilises the powerful representation learning capability of deep learning models on images to achieve effective classification of art styles.

### B. EMOTION PREDICTION FOR ARTWORK VIEWERS

In the Emotion Prediction for Artwork Viewers process, we also consider the input image as  $X$  and employ the visual

encoder  $f_{\text{encoder}}$  for feature extraction to obtain the feature map  $e$ :

$$e = f_{\text{encoder}}(X) \quad (4)$$

Similar to the Artwork Style Classification process, a pre-trained visual encoder is employed for effective feature learning. The feature map  $e$  is then mapped to the probability distribution of emotion categories through a linear layer  $g_{\text{mlp}}$  with a sigmoid activation function:

$$P(E|X) = \text{softmax}(g_{\text{mlp}}(\text{flatten}(e))) \quad (5)$$

Here,  $E$  represents the emotion category, and  $P(E|X)$  signifies the probability of emotion being  $E$  given the input image  $X$ . To measure the difference between the predicted probability distribution and the actual emotion labels, binary cross-entropy loss is utilized:

$$\begin{aligned} \text{Loss}_{\text{emotion}} \\ = \text{BCELoss}(P(E|X), \text{ground truth emotion labels}) \end{aligned} \quad (6)$$

The reason for choosing the binary cross-entropy loss function is due to the fact that emotion prediction for artwork images is a multi-label binary classification task.

For the final optimisation process we also used the Adam optimiser to update the model parameters with a combination of adaptive learning rates and weight decay.

### C. JOINT LEARNING

In the Joint Learning process, We assign style classification and emotional response analysis as Task 1 and Task 2, respectively. We first consider the input image as  $X$  and employ the pre-trained ResNet18 model  $f_{\text{ResNet18}}$  to obtain the feature vectors. For style classification:

$$\mathbf{x}_1 = f_{\text{ResNet18}}(X) \quad (7)$$

For emotion prediction:

$$\mathbf{x}_2 = f_{\text{ResNet18}}(X) \quad (8)$$

Our approach involves dedicated task-specific encoders for both tasks. For Task 1, which focuses on image style classification, the encoder architecture includes: Convolutional Layer (Conv1): This layer further processes the shared features extracted by the image encoder, capturing distinctive patterns relevant to art styles. Batch Normalization (BatchNorm1): Batch normalization is applied to stabilize training and expedite convergence, ensuring that the model effectively learns style-related features. Rectified Linear Unit (ReLU1): We incorporate the ReLU activation function to introduce non-linearity, facilitating the model's ability to capture complex style patterns. Adaptive Average Pooling (AdaptiveAvgPool1): This layer ensures a consistent feature representation size, regardless of the input image's dimensions. For Task 2, we use the same methodology and steps. The outputs for both tasks are generated through a series of layers, including linear layers and multi-layer perceptrons (MLPs). These architectural

choices enable the model to convert extracted features into task-specific representations effectively. Specifically: For Task 1, the output is generated by MLP1, providing the final image style classification result. For Task 2, MLP2 generates the output for sentiment classification. To enhance the model's robustness, regularization techniques such as dropout and batch normalization are applied within the layers responsible for output generation. These techniques mitigate overfitting and improve generalization, ensuring the model's effectiveness across diverse art images. An innovative feature of our methodology is the inclusion of a specific-task attention mechanism. This dynamic fusion ensures that the model can adaptively emphasize shared features, enhancing its ability to capture relevant information for both style and sentiment classification. This attention mechanism (in the case of Task1) first generates the attention weight  $\text{att1}$ :

$$\text{att1} = \sigma(\mathbf{W}_1 \cdot [\mathbf{x}_1; \mathbf{x}_2]) \quad (9)$$

where  $\mathbf{W}_1$  denotes the weight matrix of self.attn1,  $[\mathbf{x}_1; \mathbf{x}_2]$  denotes the splice of feature vectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , and  $\sigma$  denotes the Sigmoid activation function. By concatenating the features of Task 1 and Task 2 along the last dimension in  $[\mathbf{x}_1; \mathbf{x}_2]$ , we preserve task-specific information for each task while providing a more comprehensive and enriched representation. Multiplying the concatenated features with the trained matrix  $\mathbf{W}_1$  and combining it with the sigmoid activation contributes to the model's flexibility in learning complex relationships between tasks. Applying the attention weights to the features:

$$\text{att1} = \mathbf{x}_1 \odot \text{att1} + \mathbf{x}_2 \odot (1 - \text{att1}) \quad (10)$$

where  $\odot$  denotes element-wise multiplication.  $\text{att1}$  is the weight learned from features  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , which determines how to weight the fused features between the two tasks. This mechanism plays a vital role in fusing features from both Task 1 and Task 2, dynamically adjusting feature fusion weights. Element-wise multiplication  $\odot$  allows the model to apply different weights to the features of tasks 1 and 2 at each position, the model can dynamically adjust the focus on different position features. The term  $\mathbf{x}_2 \odot (1 - \text{att1})$  implements the opposite weighting for task 2 features, ensuring that specific information from task 1 or task 2 is not lost during fusion. This helps the model consider the contribution of each task more comprehensively.

### D. LOSS FUNCTION

#### 1) LOSS FOR STYLE CLASSIFICATION

We adopted a multi-classification approach to categorize images into different stylistic categories. In the loss function  $L_1$ ,  $C$  represents the number of categories for stylistic classification. We use  $Y_{1ij}$  to denote whether sample  $i$  belongs to category  $j$ , typically taking values 0 or 1.  $P_{1ij}$  represents the model's predicted probability that sample  $i$  belongs to category  $j$ . The loss function  $L_1$  measures the performance



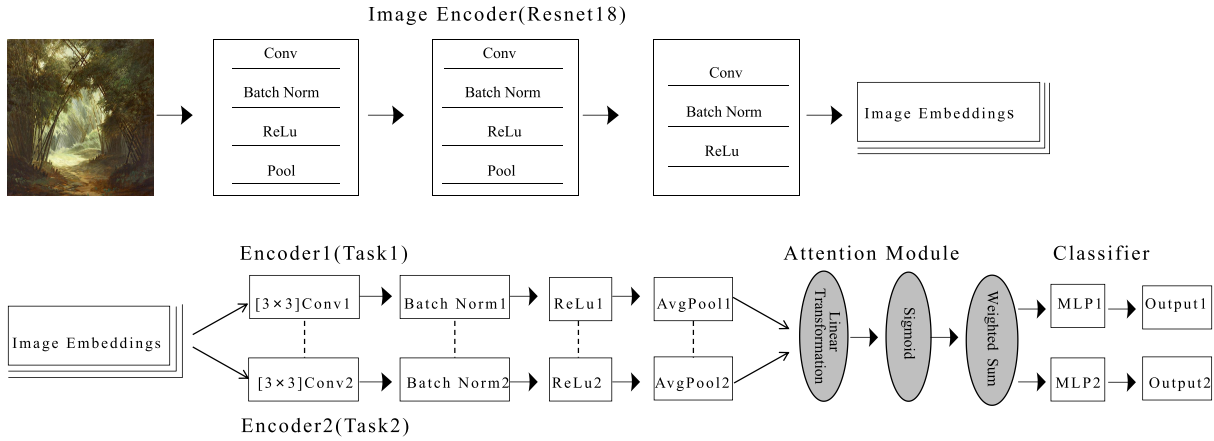


FIGURE 1. Overview of our method.

of the model in stylistic classification through cross-entropy loss.

$$L_1 = -\frac{1}{N_1} \sum_{i=1}^N \left( \sum_{j=1}^C Y_{1ij} \cdot \log(P_{1ij}) \right) \quad (11)$$

## 2) LOSS FOR EMOTION PREDICTION

We employed a multi-label binary classification method to assign multiple sentiment labels to each image. In the loss function  $L_2$ ,  $M$  stands for the number of sentiment labels.  $Y_{2ij}$  indicates whether sample  $i$  has sentiment label  $j$ , typically taking values 0 or 1.  $P_{2ij}$  represents the model’s predicted probability that sample  $i$  has sentiment label  $j$ . The loss function  $L_2$  assesses the model’s performance in sentiment classification through binary cross-entropy loss.

$$L_2 = -\frac{1}{N_1} \sum_{i=1}^N \left( \sum_{j=1}^M (Y_{2ij} \cdot \log(P_{2ij}) + (1 - Y_{2ij}) \cdot \log(1 - P_{2ij})) \right) \quad (12)$$

We obtain the output probabilities for Task 1 and Task 2:

$$P(\text{Task1output}|X) = \text{softmax}(g_{\text{mlp1}}(\text{flatten}(\text{att1}))) \quad (13)$$

$$P(\text{Task2output}|X) = \text{softmax}(g_{\text{mlp2}}(\text{flatten}(\text{att2}))) \quad (14)$$

## IV. EXPERIMENTS

### A. DATASET

Building upon the publicly available WikiArt collection, the ArtEmis dataset enriches this foundation with 80,031 carefully curated artworks from 1,119 artists, spanning 45 genres and 27 unique art styles [12]. Each artwork in the ArtEmis subset of WikiArt has been annotated by at least five annotators who have documented their primary emotional responses and provided detailed justifications for these reactions. The comprehensive collection of 454,684 emotional responses and explanatory comments in ArtEmis presents an invaluable resource for both training and evaluation purposes. The dataset’s broad coverage across

a variety of artistic styles and emotional reactions enables effective learning and inference by the model.

For our specific task, we utilize the 27 art style labels and 9 emotion categories defined within the ArtEmis dataset. The emotion labels encompass eight principal emotions and an additional category labeled “something-else,” which allows annotators to identify emotions not explicitly listed or to explain the absence of a strong emotional reaction. Furthermore, ArtEmis delves into the spectrum of emotions, uncovering attitudes, moods, and abstract concepts such as freedom and love. Annotators adeptly associate visual features with psychological assessments, shedding light on nuances in the depicted subjects. This is particularly beneficial for our analysis of emotional responses. To enhance training efficiency, all images used for training and validation have been resized to  $224 \times 224$ . This modification reduces computational complexity while preserving essential visual information. Our model is designed to predict both the art style and the emotional content from these resized images.

### B. EXPERIMENTAL SETTING

For the art style classification task using ResNet18, we train the model over 7 epochs with a batch size of 64 images. Reproducibility is ensured through a fixed random seed of 1234, and training stability is improved with a gradient clipping threshold of 5.0. These settings—7 epochs, random seed, and gradient clipping—are consistently applied across all models. The learning rate is set to  $1 \times 10^{-5}$ , and a weight decay of 0.01 is applied during optimization with the Adam optimizer. In the emotion analysis task using ResNet18, the model adheres to the same hyperparameters but with a reduced learning rate of  $5 \times 10^{-6}$ . For the style classification task employing MobileNetV2, the model is trained with a batch size of 32 images, maintaining a learning rate of  $1 \times 10^{-5}$  and a weight decay of 0.01, optimized with the Adam optimizer. For MobileNetV2 in the emotion analysis task, identical hyperparameters are used. For tasks utilizing Our version of ResNet18, the model trains with a

**TABLE 1.** Comparison of different methods on style classification.

Method	Val	Test
VGG-16	54.83	54.16
VGG-19	55.12	55.32
ViT	53.77	53.27
MobileNetv2	55.49	54.86
ResNet-18	55.47	56.23
Our (MobileNetv2)	58.27	58.53
Our (ResNet-18)	58.73	59.24

batch size of 64 images and a learning rate of  $5 \times 10^{-5}$ , with no weight decay. For tasks with Our version of MobileNetV2, the model trains with a batch size of 16 images, a learning rate of  $1 \times 10^{-5}$ , and no weight decay.

### C. ARTWORK STYLE CLASSIFICATION

We detail the outcomes of our investigation into the classification of artwork styles utilizing diverse deep learning frameworks. Our primary objective is to assess the performance enhancement offered by our innovatively proposed model, labeled as “Our.” This model uniquely incorporates both style and emotional content to elevate classification efficacy. We juxtapose the performance of our approach with several foundational models as depicted in Table 1. The assessment criteria include validation accuracy (Val) and test accuracy (Test), with the results of each model meticulously documented. Notably, Our (MobileNetv2) registers a test accuracy of 58.53%, while Our (ResNet-18) achieves the peak accuracy of 59.24%. These outcomes are superior compared to the base architectures of MobileNetv2 and ResNet-18.

**TABLE 2.** Comparison of different methods on emotion analysis.

Method	Val	Test
VGG-16	68.63	68.94
VGG-19	70.91	71.05
ViT	66.76	66.23
MobileNetv2	71.28	71.42
ResNet-18	70.93	71.06
Our (MobileNetv2)	71.51	71.67
Our (ResNet-18)	71.04	71.29

### D. ARTWORK EMOTION ANALYSIS

The results from the artwork emotion analysis illustrate a consistent improvement in performance when utilizing our proposed models, compared to traditional baseline methods. Specifically, we observe that our modified models, leveraging the MobileNetv2 and ResNet-18 architectures, demonstrate superior accuracy in emotion analysis tasks. The version of our model that incorporates the MobileNetv2 architecture achieved a remarkable accuracy of 71.67% on the test dataset, while the ResNet-18-based variant registered an accuracy of 71.29%. These outcomes highlight the effectiveness of our novel approach, which integrates the dual learning objectives

**FIGURE 2.** Attention weight analysis.

of style and emotional content. This integration facilitates a more accurate and nuanced identification of emotional expressions in artwork.

### E. DISCUSSION

The comparative analysis between Our (MobileNetV2) and standalone MobileNetV2, as well as Our (ResNet18) against ResNet18, highlights significant enhancements in performance. This affirms the added value of integrating style alongside emotional attributes into the classification schema. Such results underscore our hypothesis that a synergist approach to leveraging both stylistic and emotional elements within artwork can markedly augment classification accuracy in artwork style classification endeavors. The advancements demonstrated by our MobileNetV2 and ResNet18 based models in emotion analysis suggest that the combined learning of stylistic elements and emotional cues significantly enhances the capability for emotion recognition across different model architectures. This underscores the potential of our approach in pushing the boundaries of emotion analysis within the realm of art. Overall, the superior performance of our models in both style classification and emotion analysis underscores the effectiveness of integrating stylistic and emotional content. These findings suggest new avenues for future research, particularly in exploring more complex and nuanced models that can further enhance the classification and analysis of artistic content. The results also highlight the importance of a multidisciplinary approach, combining insights from art theory and computational techniques to achieve a more holistic understanding of artwork.

### F. ATTENTION WEIGHT ANALYSIS

In conducting the analysis of attention weights, we utilized sigmoid and mean operations on the image weights. Figure 2 (left) depicts a painting embodying the Cubism style, distinguished by its disassembly of objects into geometric forms, the presentation of multiple perspectives concurrently, and a pronounced focus on form and structure. The attention weight for the style classification (Task 1) is significantly high at 0.6130, illustrating that the model places a strong emphasis on capturing features pertinent to style during the learning process. The inherent stylized nature of Cubist art,



FIGURE 3. Case study.

with its unique amalgamation of geometric shapes and multiperspective representations, aids the model in more effectively comprehending and assimilating the distinctive characteristics of the Cubism style. Conversely, the emotion analysis (Task 2) registers at 0.4176, indicating the model's focus on the emotional content of the image. Given Cubism's focus on abstract expression through form and structure, capturing emotional content in Cubist works may pose a challenge. Here, Task 2 assumes a supplementary role, honing in on features associated with the emotions elicited by the geometric fragmentation (such as awe and amusement) and facial expressions (including anger). This focus on emotion-related features allows the model to achieve a fuller understanding of the image's content, incorporating potential emotional aspects within the Cubism style. The collaborative effort of both tasks through joint learning empowers the model to attain a more holistic comprehension of artistic images. Task 1 is dedicated to style, focusing on capturing the intricate details of Cubism, whereas Task 2 zeroes in on emotions, shedding light on potential emotional elements in Cubist works. By integrating features from both tasks, the model finds a more balanced learning approach, ensuring a thorough understanding of the image without compromising sensitivity to the nuances of Cubism style, and concurrently acknowledging the embedded emotional information.

Figure 2 (right) showcases a painting in the New Realism style. The analysis of Attention Weights yields a New Realism style Attention Weight of 0.4515 for Task 1, with the Emotional Weight slightly higher at 0.4725 for Task 2. This underscores a model preference for capturing emotionally related features in this instance. New Realism art is characterized by its precise and detailed portrayal of reality, often encapsulating realistic scenes from everyday life and urban environments, where emotional elements play a significant role. The inclusion of emotional weights enhances the comprehension of style information, effectively highlighting New Realism's focus on accurate real-life depictions. In this case, the imposing urban landscape invokes emotional responses such as contentment, awe, and excitement. Consequently, the elevated emotional weight likely mirrors the model's focus on these emotional aspects, fostering a more comprehensive grasp of the image's perceptual qualities and offering a richer emotional perspective through the model's lens.

Owing to the intricate, multi-dimensional, and rich essence of artistic images, a singular task may not suffice to encapsulate all the information within an image. The introduction of an additional task enables the model to explore and concentrate on varied facets, culminating in a more profound understanding of artistic images.

### G. CASE STUDY

After training our enhanced ResNet18 model (hereafter referred to as "Our ResNet18") alongside the standard ResNet18, we selected four exemplar images for analysis. The first image, shown in Figure 3, portrays a lion consuming its prey within a verdant jungle, encapsulating the raw essence of animal predation. The standard ResNet18 identified this image as embodying the Pointillism style, whereas Our ResNet18 accurately classified it as Naive Art Primitivism, recognizing its emotional content. This artistic style is distinguished by its simplicity, elemental nature, and ingenuous approach to representation, frequently utilizing vibrant colors, straightforward geometric shapes, and direct depictions of the natural world and life experiences [51]. The image invokes feelings of sadness, awe, and fear; sadness reflects the prey's demise, awe and fear arise from the dense jungle setting and the fundamental, instinctive act of predation. These emotions resonate with the simplicity, primordial aspect, and naïveté inherent to Naive Art Primitivism.

The second image, depicted in Figure 3, showcases a winter scene featuring a barn, interpreted by the standard ResNet18 as Realism. Conversely, Our ResNet18 discerned its Impressionist qualities, considering the emotional nuance. Impressionist art is noted for its brisk, loose brushwork that accentuates light and shadow interplay, capturing the scene's atmosphere—a stark contrast to Realism's detailed, meticulous portrayal of subjects [52]. The depicted winter landscape conveys a sense of desolation, while the solitary farmhouse amidst the chill evokes sadness. Yet, there's a sense of contentment in the harmonious coexistence of the farmhouse with its wintry surroundings, reflecting Impressionism's dynamic expression of nature and environmental moods. The alignment of conveyed emotions with the artwork enables a more precise stylistic classification.

The third image in Figure 3 illustrates the aftermath of a thunderstorm, with vivid colors and lighting denoting the potent force of nature, including elements like houses and a lake, indicative of Fauvism. The standard ResNet18 associated the depicted emotions with contentment and excitement, whereas Our ResNet18, by integrating stylistic insights, accurately recognized feelings of contentment, amusement, and excitement. The post-storm tranquility elicits contentment; the rejuvenated landscape post-deluge, excitement. Fauvism's hallmark vibrant colors and the dramatic contrast in the post-storm sky also prompt amusement.

The fourth image, presented in Figure 3, reveals willows, flowers, and a distant house, evoking an aura of peace and the beauty of nature. Where the standard ResNet18 perceived contentment alone, Our ResNet18, through stylistic integration, correctly identified feelings of contentment and awe. The peaceful willow and water scene induces contentment, while the Impressionist emphasis on lighting and natural elements arouses awe. Thus, the joint learning of style and emotional content facilitates a more accurate anticipation of the viewer's emotional response.

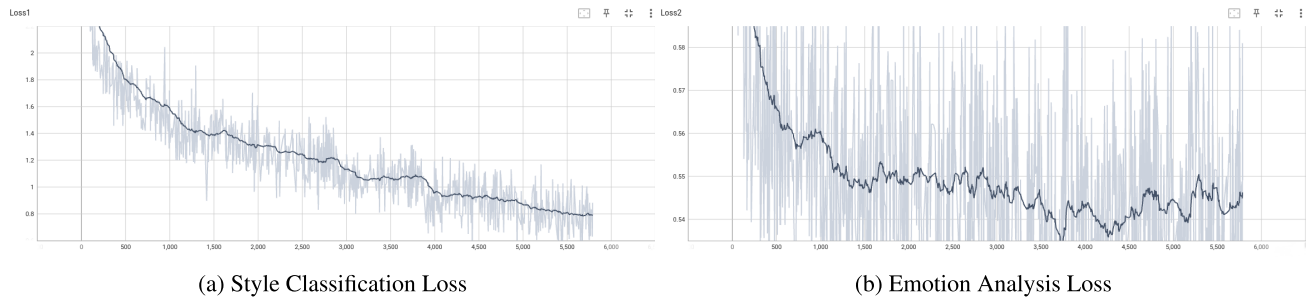


FIGURE 4. Impact of training loss.



FIGURE 5. Error analysis.

#### H. TRAINING ANALYSIS

As depicted in Figure 4, the Loss1 curve demonstrates a rapid decline from an initial value of 3.33 to 0.79 within the first 6 epochs. This rapid decrease signifies the model's swift acquisition of stylistic knowledge, indicating an effective comprehension and assimilation of style-related characteristics in artistic images. The curve's uniform descent also points to a consistent and stable learning process concerning style. Conversely, the Loss2 curve shows more pronounced fluctuations within the same timeframe, descending from 0.69 to approximately 0.54. This variability suggests that the model encounters more difficulties in learning emotional content, with challenges arising from the nuanced and complex nature of emotional information. These oscillations can be attributed to the diverse range of emotional labels present in the dataset and the marked differences in emotional content across samples. Despite these differences, the general patterns and trajectories of the Loss1 and Loss2 curves share similarities, indicating a high degree of compatibility. Both curves display a marked downward trend in the initial training phase, suggesting that the model is adept at balancing the dual objectives of learning style and emotion. This equilibrium highlights the model's capacity to effectively manage the integration of stylistic and emotional information in a unified learning task. Such a balanced approach underpins stable learning across the training period and lays a robust foundation for the model's subsequent ability to classify artistic images based on both style and emotion with relative stability.

#### I. ERROR ANALYSIS

We analyzed three artworks that were incorrectly classified by our model, ResNet18. In Figure 5 (left), the depicted painting portrays Jacob mourning for his son Joseph,

embodying the Symbolism style and expressing sadness. Our model mistakenly identified the style as Mannerism Late Renaissance and the emotion as awe. The distinction between Mannerism Late Renaissance and Symbolism primarily lies in Mannerism's focus on exaggerated, diverse forms versus Symbolism's emphasis on expressing deep emotions and meanings through symbolic elements [53], [54]. The model's incorrect prediction likely stems from its emphasis on the visual form over the intricate emotions and meanings conveyed, as Symbolism frequently employs deep symbolism and abstract expression, presenting challenges in capturing such complexity, hence leading to inaccuracies in emotion prediction. The solemn attire and collective figures might have misled the model into predicting awe instead of focusing on the central emotion of sadness portrayed by the main figure.

In Figure 5 (middle), we observe a misprediction involving a Baroque-style painting that conveys awe and excitement, related to the coronation of Maria de' Medici. The model incorrectly classified it as Early Renaissance style while correctly identifying the emotion as awe. Early Renaissance and Baroque are both Renaissance period styles with overlapping artistic elements, which may confound the model's ability to accurately distinguish between them [55]. Baroque is known for its dramatic scenes, and although the model recognized the awe emotion through the grandeur and attire, the low resolution of the image and the multitude of figures impeded a detailed analysis, thus failing to capture the excitement aspect.

Lastly, in Figure 5 (right), an artwork in the Expressionism style, evoking contentment, amusement, and excitement and titled "Landscape with Black Figure," was incorrectly classified by our model as Cubism with the emotion of amusement. Expressionism focuses on the artist's inner feelings and subjective experiences to convey emotional impact [28], whereas Cubism is characterized by the deconstruction and reassembly of geometric shapes to depict multiple perspectives [56]. The artwork's prominent geometric shapes and dimensions might have led the model to overly focus on features typical of Cubism, resulting in a misjudgment. If an artwork's stylistic features are not distinctly expressed, it can lead to inaccuracies in the model's classifications. In this instance, the model wrongly associated the artwork with



Cubism and identified amusement as the primary emotion, overlooking the broader range of emotions present. This indicates a potential model limitation in associating specific emotions with particular styles, failing to capture the diversity of emotional expressions in art. Such errors underscore the challenges models face in distinguishing between similar styles and understanding complex emotional nuances.

## V. CONCLUSION

This study has embarked on an innovative journey into the confluence of computer vision, art analysis, and emotion prediction. It introduces the Unified Model for Art Style and Emotion Prediction (ASE), showcasing a novel methodology via a multi-task learning framework that includes Artwork Style Classification, Emotion Prediction for Artwork Viewers, and a Task-Specific Attention Module. Leveraging a pre-trained image encoder alongside a task-specific attention mechanism, the architecture facilitates the concurrent processing of multiple tasks, evidencing the model's versatility and adaptability across varied artistic expressions. Our evaluation using the Artemis dataset has validated the model's adeptness at accurately classifying artistic styles and identifying emotional responses. These findings highlight the ASE model's capability to capture the subtle intricacies present within artworks, laying the groundwork for future breakthroughs at the nexus of art analysis and computer vision.

As we move into an era where the melding of art and technology gains greater significance, the contributions of this research signal a step toward a more profound comprehension of the complex interplay between visual aesthetics and human emotions. Future research may delve into refined architectures, expand upon existing datasets, and explore practical applications, further narrowing the gap between computer vision and artistic interpretation. This work, therefore, not only presents a significant advancement in understanding art through the lens of computer vision but also opens new avenues for exploration in the rich interplay between art, emotion, and technology.

## REFERENCES

- [1] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Aug. 2017, pp. 5987–5995.
- [2] H. Zhang, C. Wu, Z. Zhang, Y. Zhu, H. Lin, Z. Zhang, Y. Sun, T. He, J. Mueller, R. Manmatha, M. Li, and A. J. Smola, "ResNeSt: Split-attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Apr. 2022, pp. 2735–2745.
- [3] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in transformer," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 15908–15919.
- [4] F. Milani and P. Fraternali, "A data set and a convolutional model for iconography classification in paintings," 2020, *arXiv:2010.11697*.
- [5] N. Gonthier, Y. Gousseau, and S. Ladjal, "An analysis of the transfer learning of convolutional neural networks for artistic images," in *Proc. ICPR Int. Workshops Challenge*, vol. 12663, 2021, pp. 546–561.
- [6] J. A. Mikels, B. L. Fredrickson, G. R. Larkin, C. M. Lindberg, S. J. Maglio, and P. A. Reuter-Lorenz, "Emotional category data on images from the international affective picture system," *Behav. Res. Methods*, vol. 37, no. 4, pp. 626–630, Nov. 2005.
- [7] S. Zhao, Y. Gao, X. Jiang, H. Yao, T.-S. Chua, and X. Sun, "Exploring principles-of-art features for image emotion recognition," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 47–56, doi: [10.1145/2647868.2654930](https://doi.org/10.1145/2647868.2654930).
- [8] Q. You, J. Luo, H. Jin, and J. Yang, "Building a large scale dataset for image emotion recognition: The fine print and the benchmark," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, Feb. 2016, pp. 308–314.
- [9] Y. Zhou, T. Shen, X. Geng, G. Long, and D. Jiang, "ClarET: Pre-training a correlation-aware context-to-event transformer for event-centric generation and classification," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 2559–2575.
- [10] A. Y. Virasova, D. I. Klimov, O. E. Khromov, I. R. Gubaidullin, and V. V. Oreshko, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Nov. 2021, pp. 115–126.
- [11] Y. Zhou, X. Geng, T. Shen, G. Long, and D. Jiang, "EventBERT: A pre-trained model for event correlation reasoning," in *Proc. ACM Web Conf.*, Virtual Event, Lyon, France, Apr. 2022, pp. 850–859, doi: [10.1145/3485447.3511928](https://doi.org/10.1145/3485447.3511928).
- [12] P. Achlioptas, M. Ovsjanikov, K. Haydarov, M. Elhoseiny, and L. J. Guibas, "Artemis: Affective language for visual art," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2021, pp. 11569–11579.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9, doi: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [15] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. ECCV*, 2018, pp. 3–19.
- [16] W. Wang, X. Tan, P. Zhang, and X. Wang, "A CBAM based multiscale transformer fusion approach for remote sensing image change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6817–6825, 2022, doi: [10.1109/JSTARS.2022.3198517](https://doi.org/10.1109/JSTARS.2022.3198517).
- [17] Y. Chen, X. Zhang, W. Chen, Y. Li, and J. Wang, "Research on recognition of fly species based on improved RetinaNet and CBAM," *IEEE Access*, vol. 8, pp. 102907–102919, 2020, doi: [10.1109/ACCESS.2020.2997466](https://doi.org/10.1109/ACCESS.2020.2997466), <https://doi.org/10.1109/ACCESS.2020.2997466>.
- [18] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7794–7803.
- [19] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," in *Proc. 7th Int. Conf. Learn. Represent.*, 2019, pp. 1–12.
- [20] Z. Zhu, M. Xu, S. Bai, T. Huang, and X. Bai, "Asymmetric non-local neural networks for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 593–602, doi: [10.1109/ICCV.2019.00068](https://doi.org/10.1109/ICCV.2019.00068).
- [21] Q. Hou, L. Zhang, M. Cheng, and J. Feng, "Strip pooling: Rethinking spatial pooling for scene parsing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Sep. 2020, pp. 4002–4011.
- [22] S. Liu, L. Zhang, X. Yang, H. Su, and J. Zhu, "Query2Label: A simple transformer way to multi-label classification," 2021, *arXiv:2107.10834*.
- [23] E. Cetinic, T. Lipic, and S. Grgic, "Fine-tuning convolutional neural networks for fine art classification," *Expert Syst. Appl.*, vol. 114, pp. 107–118, Dec. 2018, doi: [10.1016/j.eswa.2018.07.026](https://doi.org/10.1016/j.eswa.2018.07.026).
- [24] X. Wu, Z. Hu, L. Sheng, and D. Xu, "StyleFormer: Real-time arbitrary style transfer via parametric style composition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, Oct. 2021, pp. 14598–14607, doi: [10.1109/ICCV48922.2021.01435](https://doi.org/10.1109/ICCV48922.2021.01435).
- [25] Y. Deng, F. Tang, W. Dong, C. Ma, X. Pan, L. Wang, and C. Xu, "StyTr<sup>2</sup>: Image style transfer with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 11316–11326, doi: [10.1109/cvpr52688.2022.01104](https://doi.org/10.1109/cvpr52688.2022.01104).
- [26] V. Tonkes and M. Sabatelli, "How well do vision transformers (VTs) transfer to the non-natural image domain? An empirical study involving art classification," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2022, pp. 234–250.
- [27] L. Schaerf, C. Popovici, and E. Postma, "Art authentication with vision transformers," 2023, *arXiv:2307.03039*.
- [28] J. E. Laird, "Emotion," in *The Soar Cognitive Architecture*. Cambridge, MA, USA: MIT Press, 2012, pp. 271–285.

- [29] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proc. 18th Int. Conf. Multimedia*, 2010, pp. 83–92.
- [30] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak, "Exploiting emoticons in sentiment analysis," in *Proc. 28th Annu. ACM Symp. Applied Computing*, 2013, pp. 703–710.
- [31] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in *Proc. 21st ACM Int. Conf. Multimedia*, Barcelona, Spain, Oct. 2013, pp. 223–232, doi: [10.1145/2502081.2502282](https://doi.org/10.1145/2502081.2502282).
- [32] S. Jindal and S. Singh, "Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning," in *Proc. Int. Conf. Inf. Process. (ICIP)*, Dec. 2015, pp. 447–451.
- [33] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 29, Feb. 2015, pp. 381–388.
- [34] G. Cai and B. Xia, "Convolutional neural networks for multimedia sentiment analysis," in *Proc. 4th CCF Conf.*, vol. 9362, 2015, pp. 159–167.
- [35] K. Song, T. Yao, Q. Ling, and T. Mei, "Boosting image sentiment analysis with visual attention," *Neurocomputing*, vol. 312, pp. 218–228, Oct. 2018, doi: [10.1016/j.neucom.2018.05.104](https://doi.org/10.1016/j.neucom.2018.05.104).
- [36] V. Lopes, A. Gaspar, L. A. Alexandre, and J. Cordeiro, "An automl-based approach to multimodal image sentiment analysis," in *Proc. Int. Joint Conf. Neural Netw.*, 2021, pp. 1–9.
- [37] Y. Ling, J. Yu, and R. Xia, "Vision-language pre-training for multimodal aspect-based sentiment analysis," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 2149–2159.
- [38] P. J. Lang, "International affective picture system (IAPS): Technical manual and affective ratings," *NIMH Center Study Emotion Attention*, vol. 1, nos. 39–58, p. 3, 1997.
- [39] D. Borth, T. Chen, R. Ji, and S.-F. Chang, "SentiBank: Large-scale ontology and classifiers for detecting sentiment and emotions in visual content," in *Proc. 21st ACM Int. Conf. Multimedia*, Oct. 2013, pp. 459–460, doi: [10.1145/2502081.2502268](https://doi.org/10.1145/2502081.2502268).
- [40] A. Dhall, A. Kaur, R. Goecke, and T. Gedeon, "EmotiW 2018: Audio-video, student engagement and group-level affect prediction," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 653–656, doi: [10.1145/3242969.3264993](https://doi.org/10.1145/3242969.3264993).
- [41] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997, doi: [10.1023/a:1007379606734](https://doi.org/10.1023/a:1007379606734).
- [42] R. Raina, A. J. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proc. 24th Int. Conf.*, vol. 227, 2007, pp. 759–766.
- [43] Y. Zhou, X. Geng, T. Shen, W. Zhang, and D. Jiang, "Improving zero-shot cross-lingual transfer for multilingual question answering over knowledge graph," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2021, pp. 5822–5834, doi: [10.18653/v1/2021.naacl-main.465](https://doi.org/10.18653/v1/2021.naacl-main.465).
- [44] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jun. 2021, doi: [10.1109/TGRS.2020.3018879](https://doi.org/10.1109/TGRS.2020.3018879), <https://doi.org/10.1109/TGRS.2020.3018879>.
- [45] A. Bensaoud and J. Kalita, "Deep multi-task learning for malware image classification," *J. Inf. Secur. Appl.*, vol. 64, Feb. 2022, Art. no. 103057, doi: [10.1016/j.jisa.2021.103057](https://doi.org/10.1016/j.jisa.2021.103057).
- [46] J. Zhao, Y. Peng, and X. He, "Attribute hierarchy based multi-task learning for fine-grained image classification," *Neurocomputing*, vol. 395, pp. 150–159, Jun. 2020, doi: [10.1016/j.neucom.2018.02.109](https://doi.org/10.1016/j.neucom.2018.02.109).
- [47] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *Proc. CVPR*, Jun. 2019, pp. 1871–1880.
- [48] Y. Liu, Z. Wang, H. Jin, and I. J. Wassell, "Multi-task adversarial network for disentangled feature learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3743–3751.
- [49] S. Chen, G. Bortsova, A. G. Juarez, G. van Tulder, and M. de Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent*, 2019, pp. 457–465.
- [50] M. Lan, J. Wang, Y. Wu, Z. Niu, and H. Wang, "Multi-task attention-based neural networks for implicit discourse relationship representation and identification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1299–1308.
- [51] R. Goldwater, *Primitivism in Modern Art*. Cambridge, MA, USA: Harvard Univ. Press, 1986.
- [52] A. Callen, *The Art of Impressionism: Painting Technique & the Making of Modernity*. New Haven, CT, USA: Yale Univ. Press, 2000.
- [53] W. B. Yeats, "Symbolism in painting," in *Essays and Introductions*. Cham, Switzerland: Springer, 1961, pp. 146–152.
- [54] D. Bodart, *Renaissance & Mannerism*. New York, NY, USA: Sterling, 2008.
- [55] H. Wölfflin, G. Ballangé, and B. Teyssèdre, *Renaissance Et Baroque*. Ithaca, NY, USA: Cornell Univ. Press, 1967.
- [56] A. Gleizes and J. Metzinger, *Cubism*. Adelphi Square, London: TF Unwin, 1913.



**CHU-ZE YIN** is currently pursuing the bachelor's degree in electronic and electrical engineering with Southwest Jiaotong University, Chengdu, China. His research interests include computer vision and control systems engineering.

• • •