IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.
Digital Object Identifier 10.1109/ACCESS.2017.DOI

AF-CPACNet : AnchorFree Crowd Parsing Attention-based Characteristic Segmentation Network

RAGHAVENDRA S[1], (Senior Member, IEEE), ABHILASH S K[2], VENU MADHAV NOOKALA.[2], and YASHWANTH NANJAPPA [3*], (Senior Member, IEEE),
[1]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, India
[2]KPIT Technologies, Bengaluru, India
[3]Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, India
*Corresponding author: Yashwanth Nanjappa (e-mail:yashwanth.n@manipal.edu).

ABSTRACT Multi-human parsing involves the task of segmenting and identifying different human parts within images that contain multiple people. It is a crucial task in computer vision, particularly for applications such as human pose estimation, scene understanding, and virtual reality. This paper explores the various features and techniques used in multi-human parsing, including the use of deep learning models like convolutional neural networks (CNNs) and attention mechanisms to accurately detect and segment human body parts in crowded or complex environments. Anchor boxes often fail to capture the diverse variations in human body shapes and poses accurately, leading to suboptimal performance in human parsing tasks. To address these limitations, we introduce AF-CPACNet, a novel model that eliminates the need for anchor boxes by adopting a multi-head and multi-task architecture. AF-CPACNet consists of two key components: a detection head and an edge-guided parsing module, enabling pixel-level analysis and improving the precision of human body part segmentation. Additionally, a refinement head is incorporated to further enhance semantic parsing quality. The model captures finer details of human body parts by considering color, size, and pattern attributes in a single forward pass while operating in real-time. A specialized loss function is employed to optimize semantic parsing results and improve training efficiency. We evaluate the performance of AF-CPACNet on multiple human parsing datasets, including CCIHP and CIHP, and demonstrate that it significantly outperforms existing state-of-the-art methods. Specifically, AF-CPACNet achieves an 11% improvement on the CIHP dataset and an mIoU of 67.3 on the CCIHP dataset, across both global and instance-level metrics. The open-source code is available at https://github.com/abhigoku10/AF-CPACNet.git

INDEX TERMS Anchor-based detection, Human parsing, Multi-task architecture, Pixel-level analysis, Semantic parsing

I. INTRODUCTION

ANALYZING humans in computer vision involve processes at both the instance and semantic levels utilizing corresponding networks. The task entails an examination of pixels to outline human figures. This challenging process holds importance as it forms the basis for various subsequent applications. Notably multi human parsing at the instance level plays a role in offering deep insights into the distinctive characteristics of each person depicted in an image. These insights are valuable for tasks like image recognition [1], pose estimation [2] and person re-identification [3], [4] as well as enhancing nuanced human-object interactions. Beyond these core uses human parsing extends to functions such as visual try-on, fashion editing [5], and fashion landmark detection.

In years there has been significant progress in human parsing, which involves identifying and labeling human body parts in images. A key focus area is instance segmentation, where the goal is to accurately pinpoint each occurrence of an object within an image. This advancement has been driven by cutting-edge learning models like FCLN with Faster RCNN [6] Cascade RCNN [7], PointRend [8] and Mask RCNN [9]. These models not only detect objects but also create detailed masks at the pixel level for each object instance. Despite strides in instance-level human parsing the most advanced method still heavily relies on the Mask RCNN model. This technique follows a two-stage process called RPRCNN [10]; first using a RCNN model to identify anchor-based proposals from specific regions and then conducting a detailed examination of human body parts, in the second stage. The Fully Convolutional Localization Network (FCLN) combines object localization and image captioning in a single pass, achieving superior accuracy and contextual understanding, especially in complex scenes [6].

The approach of using a two-stage anchor-based detector has certain limitations. One of the drawbacks of this approach is that it cannot provide pixel-wise predictions. This can have an impact on the accuracy of per-pixel parsing, especially for tasks that require instance-level precision. In such tasks, the objective is to label each pixel accurately based on the specific instance of a body part it belongs to. Although there have been significant advancements in human parsing, particularly in instance segmentation, the current state-of-the-art approach still has room for improvement in terms of consistency and precision.

Currently, a “Fully Convolutional Network-based Detector” (FCOS) [11] is being developed as a novel technique. This approach is gaining attention due to its potential to deliver more precise human recognition results through its exhaustive per-pixel representation. It is believed to be a promising alternative to traditional methodologies. In order to attain accurate instance parsing regions with comprehensive properties, edge-guided parsing integration is important. This is achieved by combining the human parsing branch with a simple edge prediction branch. The method greatly enhances parsing results by utilizing part-level information and box-level scores along with part-aware criteria like mIoU loss. At the heart of this significant shift in thinking is a new type of refinement head. This head has been meticulously designed to improve parsing results throughout a variety of part and global-level metrics. The refining head takes into account color, size, and pattern parsing attributes in addition to edge-guided parsing. It has a more precise attention network, which improves segmentation results all around. These enhancements demonstrate the effectiveness of crowd parsing techniques at the production level while also lowering the computing costs related to comprehensive analysis within multi-human parsing networks. Overall, these advancements enable transformative growth and enhanced applicability of human parsing across diverse domains.

The AnchorFree Crowd Parsing Attention-based Characteristic Segmentation Network was developed to improve the efficiency and accuracy of parsing crowd scenes, which pose challenges due to the high density of individuals with diverse poses and orientations. Conventional segmentation methods struggle to segment crowded scenes limiting their applications in crowd monitoring, urban planning and public safety.

In this study, our proposed architecture extracts multiple human parsing features, including color, size, and pattern, in a single forward pass while functioning in real-time. In contrast, previous studies have generally focused on extracting only one parsing feature at a time and were not suitable for real-time applications. AF CPACNet has advanced computer vision by addressing the challenges of instance segmentation in crowded scenes. It enhances system’s ability to identify individual instances within crowded images aiding in the interpretation of complex visual environments. This work’s primary contribution lies in its advancements towards improving instance segmentation in challenging scenarios as follows.

- A new multi-task network named AF CPACNet has been designed for a variety of human parsing applications. It can identify body parts and instance-level features using attributes providing an approach to analyzing humans.
- Incorporating an attention network into AF CPACNet enhances the extraction of features for human-centric analysis. This integration allows the model to gather information from multiple individuals in a scene resulting in more accurate parsing outcomes.
- The research conducts experiments on the CCIHP dataset to evaluate body parts, sizes, patterns and color maps. Segmenting these attributes is crucial for achieving segmentation results in human parsing tasks.
- By introducing a customized Intersection over Union (cIOU) loss function the study achieves accuracy on the CCIHP dataset. This unique strategy improves the models adaptability and facilitates integration showcasing its compatibility and detailed analysis capabilities.
- The proposed multi-task framework with the cIOU loss function enhances the model’s flexibility and resilience efficiently handling parsing tasks while maintaining high levels of accuracy. This advancement represents progress, in human parsing.

The paper is divided into parts, each of which focuses on a specific aspect of the research. In Section II an overview of Related Works is provided, discussing studies conducted in the field. Section III details the Proposed Methodology introducing an approach to the study. The datasets used and the experimental setups are described in Section IV. Results from the experiments are presented in Section V along with a discussion, of these outcomes. A detailed analysis of ablation studies to assess components of the proposed method is covered in Section VI. Lastly, Section VII wraps up the paper by summarizing discoveries and suggesting future research directions.

II. LITERATURE SURVEY

The AF CPACNet is a network system comprising multiple interconnected modules. In this segment, we will examine the literature and recent research that investigates the various elements of this network. Our aim is to enhance our comprehension of how these modules collaborate to accomplish the systems overarching objectives.

A. SEGMENTATION

In computer vision, segmentation is an essential procedure that divides an image into several segments or regions according to predetermined standards. One of the essential steps in segmentation is instance-level object segmentation, which involves identifying and predicting the precise boundaries of each object in an image by creating a full mask for it. This technique enables computers to recognize and differentiate between different objects in an image accurately, which is essential for various applications, including object detection, autonomous driving, and image annotation.

Human Parsing Some researchers combine human body parts into a universal representation in order to produce powerful representations. Li et al. [12] identified the major body parts—such as the torso, arms, legs, etc.—using a spatial transformer network. To improve the localization of body parts a region proposal network was used by Zhao et al. [13] that was trained and tested with the help of an auxiliary dataset. Mahdi et al. [14] attempted to use local cues to address person re-identification. Several patches representing different human body parts were retrieved, yet part misalignment could not be adequately addressed by the technique. To overcome this limitation, the researchers proposed integrating pose estimation into the person re-id process. They suggested using conventional pose estimation models to initialize part locations and subsequently aligning them using an affine transformation or spatial transformer network.

Characteristic Segmentation Computer vision encounters difficulties when trying to distinguish humans based on patterns, size, and color. There are potential applications for this task, such as human identification, surveillance, and fashion analysis. Liu et al. [15] introduced a method for segmentation that utilizes a depth-enhancing attention network. This approach combines color and depth images to predict body shape and skin tone, which is then used in a graph-based technique for segmenting the body. Zhao et al. [16] have devised a solution for analyzing crowded scenes with multiple human subjects by employing a deep nested adversarial learning framework that merges deep learning methods with adversarial training techniques. This framework aims to detect anchor points that are used to partition the human body into different scales enhancing the accuracy of identifying multiple humans in crowded settings. The performance of this approach has been evaluated across advanced architectures and has produced competitive outcomes. In another study [17] titled "AIParsing; Anchor instance level human parsing," a novel method for human parsing, using anchor-free instance level strategies is presented. This approach aims to enhance the accuracy and efficiency of human parsing, which is crucial for various computer vision tasks such as human-computer interaction, action recognition, and virtual try-on. The existing anchor-based methods limitations are addressed by the framework which achieves state-of-the-art performance in human parsing. The architecture of AIParsing has been benchmarked on one of the best human segmentation datasets called CCIHP using anchor-free attention networks and has achieved an mIOU of 79%. This paper demonstrates several ways for human segmentation on the basis of color, size, and patterns, combining both contemporary deep learning-based techniques and conventional statistical methods.

B. ATTENTION NETWORKS

In deep learning models, attention networks have become increasingly popular for computer vision tasks involving detection and segmentation. In recent years, deep learning detection and segmentation techniques have increasingly integrated attention networks. Kaiming et al. [9] introduced a region-based convolutional neural network that incorporates an attention mechanism along with a mask branch. The outcomes showed potential for future applications, but they were limited in their ability to handle unpredictable scenarios. However, a two-stage attention network developed by Dongli et al. [18] has been shown to produce highly accurate results for semantic segmentation in natural scenes. . To improve performance, this network makes use of both spatial and channel attention methods. Self-attention feature fusion network (SA-FFN), developed by Zhao et al. [19], enhances semantic segmentation task performance. The network consists of multiple encoder and decoder modules, each equipped with self-attention mechanisms that enable the network to focus on informative regions while suppressing irrelevant information. The experimental results on datasets such as PASCAL VOC, Cityscapes, and ADE20K demonstrate that SA-FFN delivers remarkable performance in semantic segmentation tasks.

C. MULTI TASK LEARNING FRAMEWORKS

Michael et al. [20] have provided an extensive review of multi-task learning (MTL) techniques using deep neural networks (DNNs). They have covered the fundamental concepts, recent advancements, and applications of MTL in various domains. The motivation behind MTL includes improving generalization, model efficiency, and feature learning. The survey covers a wide range of topics, including different MTL architectures, optimization strategies, regularization techniques, and evaluation metrics. Wei et al. [21] have introduced a novel approach to enhance the identification of cancer driver genes. The model integrates various genomic data types and captures complex gene interactions across multiple tasks using multi-task learning and graph convolutional networks. It outperforms current methods in actual cancer datasets, indicating the potential for advancing cancer research and personalized medicine. The computational complexity of using multi-task learning on graph convolutional networks may limit scalability and practical applicability, especially for large-scale genomic datasets. In the AF-CPACNet architecture, MTL has been effectively integrated, surpassing the limitations of previous MTL frameworks by leveraging logit gradients from convolutional layers to reduce

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3501394

computational load while maintaining accuracy, particularly through the use of dual-stage networks.

D. VARIOUS APPLICATIONS OF CNN

In [1], Facial recognition explores two detection approaches: a CNN with KNN classification and a Siamese network. This study introduces a novel encryption method for color images that enhances the Vigenere cipher with chaotic maps and dynamic pseudorandom affine functions, showing significant resilience against cryptographic attack [22] This study introduces WGH-net, a CNN-based underwater image enhancement network that effectively addresses issues like color casts and low visibility in deep ocean images [23]. The MS-Xception model is an enhanced CNN-based method for grape leaf disease classification that outperforms ResNet50 and improves diagnostic efficiency in grape production [24]. This paper presents a novel global image segmentation model that excels in accurately segmenting complex images with intensity inhomogeneity and noise by employing advanced techniques like active contour methods and local similarity factors. [25].

III. PROPOSED METHODOLOGY

In the proposed model a task learning (MTL) setup incorporates two predictor heads [26], [27], within the anchor free network to improve segmentation and detection. Figure 1 and Figure 2 illustrates AF-CPACNet architecture design and the methodology section elaborates on each component of the architecture.

A. BASELINE

In the initial stage, the input image is processed with backbone ResNet50 and scaled to a fixed size of 256x256. The mean RGB values are subtracted from each pixel of the image in the dataset to normalize the value and divide the pixel value from the standard deviation. Multiple convolutional layers, followed by batch normalization and ReLU activation functions are utilized in the model. The fixed-length feature vector from the feature map is generated after completing the backbone process. The fixed-length feature vector is then processed to significantly improve the feature representation. The "Feature Pyramid Network" (FPN) or the "Global Semantic Enhancing Module" (GSE-FPN) modules are used to refine and classify to improve the results.

B. BI-DIRECTIONAL FPN BLOCK

The most commonly used method for object detection is the Bi-directional Feature Pyramid Network (BiFPN). The input image is passed through the FPN-based backbone to gather multi-level features, enhancing the BiFPN process. P2, P3, P4, P5, and P6 represent the five semantic levels of feature mapping for different strides contained in BiFPN. A single convolution layer with top-down approach generates P2, P3, and P4 from the feature maps C2, C3, and C4 of the Convolutional Neural Network (CNN). P4 and P5 represent the two strides applied by the convolution layer to produce P5 and P6, respectively. Utilizing the features of BiFPN to analyze features across two subnetworks and compute human bounding boxes.

C. ATTENTION NETWORK

To capture characteristics effectively the AF CPACNet incorporates an attention network with global average pooling in its BiFPN layers. Once the feature maps are created they undergo processing in an average pooling layer to consolidate values into single-channel data. This consolidated output is then directed to a connected layer responsible for generating the AF CPACNet detection head. The attention mechanism prioritizes regions of the input image based on learned weights adjusting network activations accordingly during training. Following feature map creation the next stage involves their transfer to an average pooling layer before routing to a fully connected layer, for AF CPACNet detection head generation. By leveraging an attention mechanism the network can focus on image areas while global average pooling streamlines feature map dimensions and produces a more succinct representation. These strategies aim to boost network performance while reducing complexity.

D. FCOS HEAD

This head portion is based on FCOS [11] and features an anchor-free detector that utilizes full convolutional processes to avoid parameter sensitivity. The FCOS head has three main branches: classification, regression, and centerness. The probability that an object will be detected at each frame is predicted by the classification branch. The regression branch predicts how far the projected bounding box will shift from the anchor box's center. The centerness branch predicts the anticipated bounding box's centerness score, a measure of the predicted box's proximity to the object's center. Each point in the output feature map corresponds to a position in the input image, and the classification and regression branches are implemented as fully convolutional networks. After applying non-maximum suppression to eliminate redundant detections and thresholding the classification and centerness scores, the final set of detections is then obtained.

E. SEMANTIC EDGE REFINED BLOCK

The semantic edge parsing head comprises the PGEC module and the refinement head for parsing, along with its different characteristic segmentation.

PGEC Module: : In order to attain more comprehensive feature characteristics and improved segmentation of human instances, finer resolutions are preferred for a specific objective. However, using a coarser resolution to achieve a high level of information is possible by utilizing the “Pyramidical Gather- Excite Context module” (PGEC). This is accomplished by establishing connections between several spatial locations in an image, where a human part is represented by each position name. By connecting the respective spatial positions and contexts of various body parts, the relationship between them can be represented.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3501394

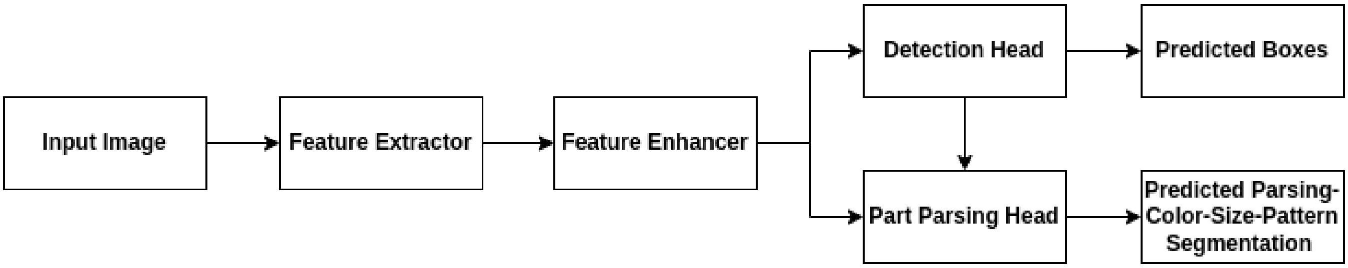

FIGURE 1. Flowchart of the proposed AF-CPACNet architecture

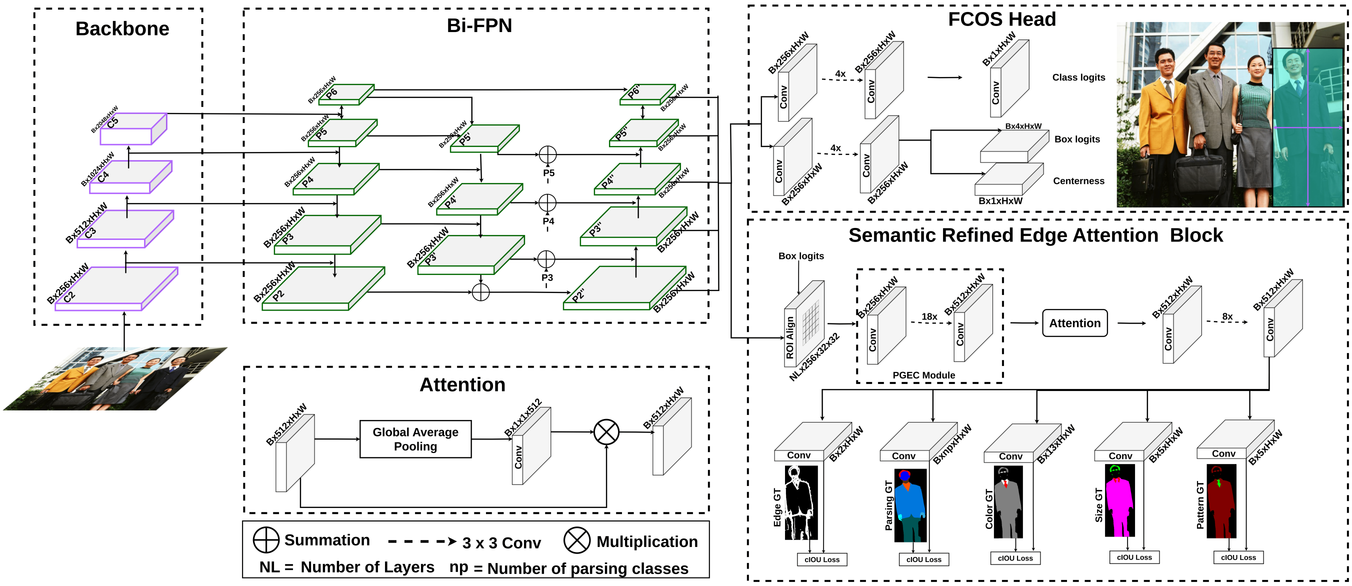

FIGURE 2. The proposed AF-CPACNet architecture includes a ResNet50 Backbone, Bi-FPN, ECA-Net, FCOS network, and a parsing refinement head for attribute parsing. The cIOU loss function is used for evaluation.

Refinement Head: The Refinement Head is responsible for two tasks: making edge parsing predictions and enhancing them. Segmenting adjoining human body parts helps to achieve more accurate segmentation of the human body. The proposed AF-CPACNet predicts more accurate regions by utilizing edge information, compared to Parsing R-CNN and AIParsing. To make the distinctions between overlapping human cases obvious, edge cues are also used. Segmentation predicts accurate mask boundaries using edge clues. By differentiating between various human instances, AF-CPACNet enables more accurate instance region localization.

F. LOSS FUNCTION

Detection Loss Both localization and classification are essential for object detection. In computer vision, bounding boxes are widely used to identify and locate multiple objects in an image. The process of detecting these bounding boxes and their corresponding regression classification values is commonly achieved by using the Smooth L1 loss function. The sizes and coordinates of the bounding boxes can be accurately predicted with the use of this statistical technique, enabling object detection and recognition in various applications.

$$L_{det} = \begin{cases} 0.5(a-b)^2 & if|a-b| < 1 \\ |a-b| - (0.5) & otherwise \end{cases} \tag{1}$$

Where

- The predicted BBox loss is represented as 'a'.
- The actual values are represented as 'b'.
- Its sensitivity to outliers and prevents gradient explosions by using loss function reduces.

Parsing Loss The difference between each human instance's expected result and the predicted segmentation map is known as the regression target error. This allows for a comparison between the segmentation projection for each instance and the ground truth segmentation map to determine its reliability. The parsing loss equation, represented as L_{parse}, is given below:

$$L_{parse} = -(b \log(x) + (1-b)\log(1-x)) \tag{2}$$

Characteristic Loss The cIOU loss, known as " Complete Intersection Over Union " serves as a measure in object detection tasks to assess the accuracy of predicted bounding boxes by determining the overlap between ground truth and anticipated bounding boxes. It calculates the ratio of the

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3501394

Category	Architecture	Backbone	Input Size	Params (M)	Flops (G)	CIHP			
						Box metrics	Parsing Metrics		
						mAP	mIOU	APpvol	PCP50
Body Parts [Bottom Up]	DeepLab v3+ [28]	Xception	299 x 299	-	-	-	58.9	-	-
	PCNet [29]	ResNet101	512 x 512	55	250	-	61.1	-	-
Body Parts [Two Stage Top Down]	SCHP+ [30]	ResNet101	473 x 473	67	78	-	67.5	-	-
Body Parts [One Stage Top Down]	ParsingRCNN [31]	ResNet50	512 x 1400	53.5	520.06	71.22	58.75	54.66	60.7
	RPRCNN [10]	ResNet50	512 x 1400	57.59	740.53	70.9	62.51	60.12	64.9
	QAnet [32]	ResNet50	384 x 512	31.47	43.15	-	64.65	59.7	68.9
	AIParsing [31]	ResNet50	640 x 800	48.77	469.71	72.7	60.88	57.79	66.3
	HPTR [33]	Resnet50	512 x 512	-	-	-	53	-	
	Ours	Resnet50	640 x 800	80.55	870.6	**75.9**	**65.4**	**64.1**	**70.2**

TABLE 1. The table clearly compares the proposed AF-CPACNet with state-of-the-art (SOTA) architectures [26] on the CIHP dataset. AF-CPACNet is rigorously evaluated using global-level parsing metrics, including mIOU, AP_pvol, and PCP50, to demonstrate its effectiveness in segmentation.

intersection of these boxes to the sum of predicted and actual bounding boxes. This metric is widely applied in learning models to enhance object detection precision and localization.

The loss function plays a role in semantic feature segmentation for recognizing the characteristics of an object. Equation 3 presents an expression of this loss function critical for identifying an object's semantic attributes. Moreover, in scenarios involving class labeling setting M to 2 in the equation results, in an improved attribute classification loss function enabling more precise and efficient classification of the attributes of an object.

$$L_{cmap} = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{3}$$

Total Loss AF-CPACNet is a multitasking framework that uses a single loss function for segmentation and parsing. Ground truth and prediction labels are used to compute the loss on a per-pixel basis. The total loss for AF-CPACNet is defined as the summation of segmentation loss (L_parse), detection loss (L_det) and characteristic attribute head loss (L_cIOU).

$$L_{Total} = L_{det} + L_{parse} + L_{cIOU} \tag{4}$$

IV. EXPERIMENTAL PROCEDURES

A. DATASETS

The CIHP dataset is a thorough collection of images featuring individuals in various settings like public areas and offices. It comprises more than 50,000 images each meticulously annotated with detailed information about every individual in the picture including body parts and clothing characteristics. This dataset is segmented into training, testing and validation sets in an 80;10;10 ratio making it highly versatile for applications. The CIHP dataset finds utility across domains such as human-computer interaction, surveillance and fashion analysis due to its rich annotations that offer substantial value to researchers and developers.

In the realm of multi-person parsing there is currently no dataset that provides localized features along with attributes such as size, color and pattern. Noteworthy advancements can be observed in the CCIHP dataset which stands as the most extensive multi-HP dataset available. Constructed from the CIHP dataset this new compilation consists of 5,000 validation and test sets along with 28,280 training samples. It encompasses 20 pixel annotated classes along with 19 semantic attribute classes. The dataset covers 110,821 annotated individuals, with masks depicting clothing items and accessories without specific labels. Typically there are three individuals in each picture with more than 20 characteristics, like glasses, skirt, dress, hair and facial hair. In the training phase, all experiments are carried out using the Tesla V100 GPU and the PyTorch framework. Our model leverages the ResNet50 architecture with a batch size set at 16. Specifically, in the Bbox branch, every image comprises 512 sampled Regions of Interest (RoIs), whereas the Parsing branch utilizes only 16 sampled RoIs. To enhance the model's flexibility and prevent overfitting, the proposed method applies augmentations such as ColorJitter, RandomHorizontalFlip, RandomCrop, and other conventional techniques. When conducting training on the CIHP dataset, the process involves a total of 200,000 training iterations, each with a reduced learning rate set at 1×10^{-3}.

V. RESULTS & DISCUSSION

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3501394

CCIHP Dataset					
Domain	Metrics	Category	Architecture		Improvement
			HPTR [33]	Ours	by AF-CPACNet
Detection Box	mAP	BBox	-	76.8	
Characteristic Parsing	mIOU	(Color)	43.8	49.75	**+5.0**
		(Size)	58.8	67.2	**+8.4**
		(Pattern)	67.4	70.2	**+2.8**
		(Body Parsing)	52.1	67.3	**+15.2**
	APrvol	(Color)	15	23.5	**+8.5**
		(Size)	24.5	28.07	**+3.57**
		(Pattern)	20.9	24.7	**+3.8**
		(Body Parsing)	29.7	37	**+7.3**
	PCP50	(Color)	-	51.7	-
		(Size)	-	49.2	
		(Pattern)	-	52.3	
		(Body Parsing)	-	71.2	

TABLE 2. Performance Comparison of AF-CPACNet Against State-of-the-Art Architectures on the CCIHP Dataset Across Different Attributes (Color, Size, and Pattern).

A. COMPARISON WITH SOTA

AF-CPACNet's efficacy was validated by analyzing its performance in comparison to state-of-the-art bottom-up and top-down parsing approaches on CIHP and CCIHP validation sets, which are shown in Table 1 and Table 2 respectively.

CIHP The AF-CPACNet was tested on the CIHP dataset using state-of-the-art architectures such as DeepLab 3+ [35]. The model achieved an mIOU score of 58.9 when evaluated on the dataset. The model incorporates multi-scale contextual data by implementing a modified atrous spatial pyramid pooling (ASPP) module with an Xception backbone. This approach is complemented by a global pooling branch, which effectively captures the overall context of the image. By combining these features, the model is able to analyze the image data at different scales and contexts, leading to more comprehensive and accurate understanding of the visual information.

In the category of bottom-up methods, promising results were shown for the PCNet model [36]. This model analyzes each region to determine the presence and location of various body parts by gradually segmenting the image into smaller regions. The key advantage of PCNet is its capability to handle images with multiple persons and complex body poses, yielding an output of 61.1 mIOU score. However, due to a lack of datasets and annotation errors, SCHP+ [37] proposed a new model that corrects the estimate in real-time while training and converges the output using the revised estimate to provide a new and more accurate pose estimate. When evaluated on the CIHP dataset, it was able to concurrently achieve an mIOU of 67.5, which is relatively higher than the models using the bottom-up approach.

For instance level parsing metric The AF-CPACNet has demonstrated superior performance compared to all state-of-the-art models, achieving an impressive mean Intersection over Union (mIOU) score of 65.4 using the one-stage-top-down approach. It's worth noting that these approaches have been known to be challenging to scale and have computational times that depend on the scene's density. RP R-CNN typically processes CIHP images in 136 to 195 milliseconds, depending on the number of persons per image (between 2 and 18). Even though Parsing R-CNN utilizes the COCO dataset to enhance its parsing performance, AIParsing outperforms it because it does not rely on COCO pre-training. AIParsing uses test-time augmentation strategies to achieve a remarkable mIoU score of 60.88%. Because of its "Quality-Aware Module" (QAM), which integrates numerous quality information, and its pixel score based on a probabilistic map, QAnet [19] gets an even higher mIoU score of 64.65%. The final global parsing results are obtained in the inference step by combining the outputs from one detection head and two parsing models. The average precision value is 64.1 and 70.2 for the instance-level parsing body metric PCP. Consequently, At the semantic and instance levels, the proposed AF-CPACNet achieves higher metric scores and shows improved human parsing performance.

FIGURE 3. Visual Representation of Color Characteristic parsing of the AF-CPACNet on the CCIHP dataset with the ground truth (GT) annotations.

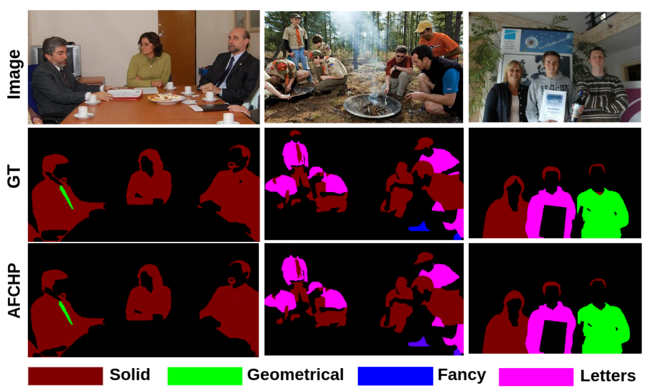

FIGURE 4. Output illustration of the proposed AF-CPACNet on different pattern attributes of the CCIHP dataset.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3501394

Attribute -Body Parts	CIHP						CCIHP			
	RPRCNN [10]	ParsingRCNN [31]	QANet [32]	AIParsing [34]	HPTR [33]	Ours	HPTR [33]		Ours	
	mIOU						mIOU	APvol	mIOU	APvol
Hat	54.6	50.2	58.6	56.3	58.1	59.7	55	52	60.9	57.9
Hair	71.1	69.3	76.3	70.9	70.4	78.1	69.7	60.9	77.52	63.54
Glove	21.8	15.7	16.6	16.9	10.5	25.3	7.5	6.6	25.41	15.61
Sunglasses	28.8	26.1	28.6	29.7	33.9	34.1	-	-	-	-
Glasses	-	-	-	-	-	-	41	32.1	51.98	35.4
UpperClothes	59.3	55.9	62	58.3	53.5	63.44	48.9	45.1	62.3	48.41
Dress	42.9	35.4	37.4	39.1	34.6	46.9	-	-	-	-
Mask	-	-	-	-	-	-	8.2	10.7	27.9	26.1
Coat	49.4	43.9	48.6	47.7	41.2	51.5	44.7	48.3	49.89	50.65
Socks	20.4	18.2	21.9	20.1	17.3	23.6	16.8	12.8	22.54	16.87
Pants	55.1	52.9	60.9	54.5	49.2	59.14	50	52.8	58.99	57.59
Torso-skin	59.1	57.1	65.7	60	57.2	62.3	57.8	48.1	65.58	53.12
Scarf	25.1	20.3	18.4	19.1	7.6	30.5	-	-	-	-
Scarf/Tie	-	-	-	-	-	-	34.1	28	46.1	31.82
Skirt	28.1	23	25.9	23.9	19.5	32.1	38	40.3	32.65	35.41
Face	71.8	70.9	79.6	72.9	69	76.2	66	71.3	77.98	75.19
L-arm	62.7	58.7	68.4	62.2	38.9	62.1	38.8	9	62.11	15.4
R-arm	62.4	58	69.2	61.6	31.7	66.45	30.2	11.6	64.03	23.6
L-leg	48.7	43.6	50.7	48.3	24.1	51.87	23.7	10.8	51.1	33.6
R-Leg	45.7	41.4	48.5	45.3	18.8	49.56	19.5	8.4	47.8	18.05
L-shoe	38.1	31.1	44.3	39.9	20.1	44.3	19.5	7.2	44.68	20.8
R-shoe	36.33	31.9	44.3	39.9	16.9	42.23	17.1	7.2	41.5	20.74
All	62.51	58.8	64.7	60.9	53	**65.4**	52.1	29.7	**67.3**	**37**

TABLE 3. Quantitative Analysis of the proposed AF-CPACNet architecture with the existing SOTA architecture [26] based solely on the Body Parsing Attributes evaluated on both CIHP and CCIHP in this research work.

B. QUANTITATIVE ANALYSIS

In this section, we aim to validate the proposed architecture by utilizing the CCIHP dataset, which stands out for its comprehensive mapping of additional attributes such as color, size, and pattern-level segmentation. By incorporating these attributes alongside body parts, the dataset offers a rich source of information for human analysis within a specific frame or image. The qualitative analysis of AF-CPACNet can be observed in Figure 6.

Analysis on Human Parsing: The proposed approach can bypass the detection gap between the "state-of-the-art" (SOTA) one-stage top-down techniques because to its lower latency and quicker computations [5]. One reason for the consistency between characteristic and attribute masks in HPTR is that task branches share backbone features. AF-CPACNet outperforms HPTR considerably and obtains SOTA outcomes with the same baseline. In comparison to HPTR, it leads to complete suppression in all parsing metrics, both semantic and instance-level. For both APp vol and PCP50, the AF-CPACNet that uses single-scale testing produces better results. The attribute-wise analysis of AF-CPACNet is provided in Table 3. It is observed that AF-CPACNet is able to achieve higher mIOU for various CCIHP attributes like 'glasses', 'skirt', 'R-arm' and 'glove'. The overall mIOU of the proposed network is 64.2% and achieves 33.7 APvol, which is comparatively greater than the HPTR architecture on the CCIHP dataset.

Color Characteristics The CCIHP dataset was used to evaluate semantic color parsing, and detailed information obtained using the additional head is presented in Table 4 and shown in Figure 3. It has been observed that the AF-CPACNet performs well in terms of the global metric mIOU and APvol for almost all colors. Particularly for the semantic colors 'Brown', 'Red', 'Yellow', 'Orange', and 'Purple', the AF-CPACNet achieves exceptional scores, setting the benchmark for the CCIHP dataset with an overall mIOU of

FIGURE 5. Visual representation of AF-CPACNet on size attributes for CCIHP dataset

49.75% and APvol of 23.5.

Pattern Characteristic The four pattern categories - "Geometrical", “Solid”, "Letters" and "Fancy," - describe the characteristics of clothing and hair. The "Solid" category contains all "Hair" labels and a large portion of the clothing, making it three times more prevalent than the other categories. The proposed AF-CPACNet achieves better scores in all categories, with an overall mIOU score of 70.2 and an AP score of 24.7 as illustrated in Figure 4 and Table 6.

Size Characteristic: Four size classes are used to categorize clothes and/or hair characteristics: "Short/small," "Long/large," "Undetermined," and "Sparse/bald" (for the "Hair" class only). Attribute classes and the size characteristic are combined to produce detailed attribute labels. For example, "Pants" + "Short" means shorts, and "Shoes" + "Long" equals boots. An mIOU score of 67.2 and an AP score of 28.07 were achieved when compared to the HPTR model benchmarked on the CCIHP dataset, demonstrating that AF-CPACNet has outperformed the state-of-the-art model, as shown in Figure 5 and Table 5.

Color	HPTR [33]		Ours	
	mIOU	APvol	mIOU	APvol
Dark	60.3	40.8	62.40	41.23
Medium	23.1	18.4	28.80	33.6
Light	31.5	25.2	38.93	33.7
Brown	9	5.5	19.63	17.6
Red	19.7	17.4	27.24	24.32
Pink	11.4	9.1	17.01	13.97
Yellow	11.2	9.6	19.51	17.58
Orange	1.1	0.07	22.26	19.35
Green	12.4	14.8	21.87	20.1
Blue	15.7	22.3	21.95	29.6
Purple	4.9	3.2	15.96	13.7
Multi-Color	14.2	13.2	17.61	17.4
All	**43.8**	**15**	**49.75**	**23.5**

TABLE 4. Quantitative Analysis of the Proposed AF-CPACNet for Color Characteristic Parsing on the CCIHP Dataset.

Size	HPTR [33]		Ours	
	mIOU	APvol	mIOU	APvol
Short	55	33.1	58.02	37.8
Long	58.6	37.5	61.69	41.1
Undetermined	20.5	13.5	21.22	17.8
Sparse (bald)	14.6	13.7	18.76	15.6
All	58.8	24.5	**67.2**	**28.07**

TABLE 5. Quantitative Comparison of AF-CPACNet for Size Attribute on the CCIHP Dataset

Pattern	HPTR [33]		Ours	
	mIOU	APvol	mIOU	APvol
Solid	72.3	36.9	76.13	41.6
Geometrica	30	14.4	35.90	18.03
Fancy	21	14.1	21.28	17.95
Letters	16.3	18.2	20.79	21.1
All	**67.4**	**20.9**	**70.2**	**24.7**

TABLE 6. performed the Quantitative Analysis of the proposed AF-CPACNet on Pattern Characteristic parsing evaluated on CCIHP dataset.

VI. ABLATION STUDIES

In this section, the effectiveness of AF-CPACNet for each part is analyzed using ResNet50 as the baseline on the CIHP and CCIHP validation datasets. The results are shown in Table 7 and Table 8.

Analysis on Body Parts The first validate AF-CPACNet using only body parts (BP), achieving a global mIOU metric of 64.9 for CIHP and 66.6 for CCIHP datasets.

Analysis on Body Parts and Color Mapping Introducing semantic color mapping (CM) alone onto AF-CPACNet results in an average mIOU score of 47.4 on the CCIHP dataset. In the suggested multi-task framework, combining BP and CM results in a considerable change in mIOU scores.

Analysis on Size and Pattern Characteristics Adding size mapping (SM) along with BP and CM does not significantly improve mIOU scores for both datasets. On the other hand, adding Pattern mapping (PM) to the MTL framework results in a 3

Inclusive Augmentation The experiment involves performing augmentations to enhance the flexibility of AF-

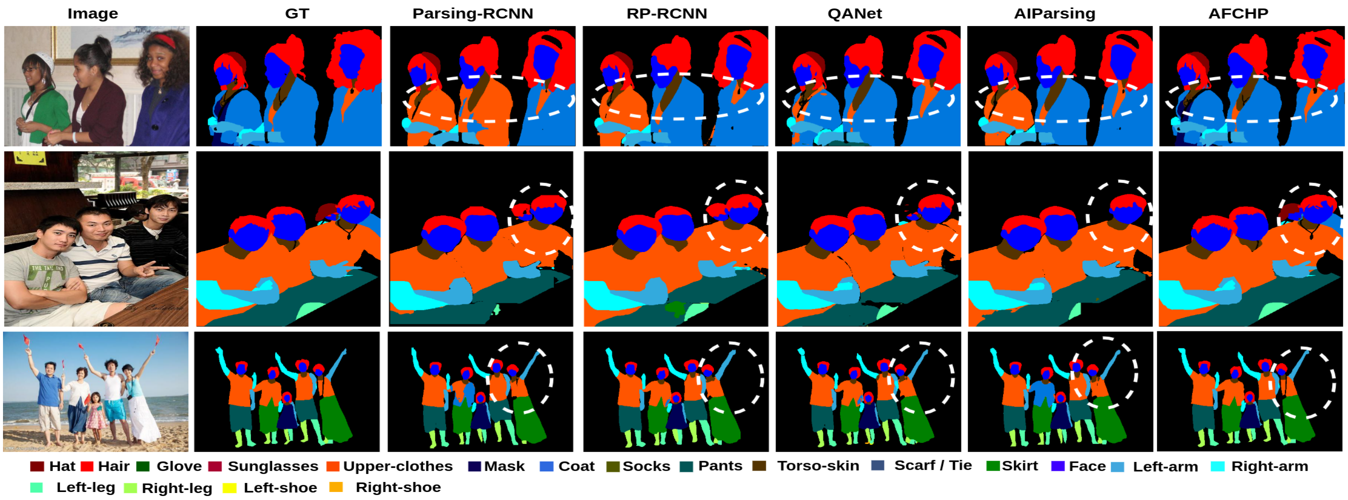

FIGURE 6. Qualitative Analysis of the Proposed AF-CPACNet Architecture Compared to Existing State-of-the-Art Architectures on the CCIHP Dataset.

(a) Input Image (b) Inaccurate Segmentation

FIGURE 9. The AF-CPACNet approach for human parsing encounters limitations, including mask bleeding in densely crowded scenes due to overlapping receptive fields in feature maps, leading to inaccurate segmentation.

							CIHP		CCIHP				
Backbone	Parts parsing	Color parsing	Size parsing	pattern parsing	Augmentation	Attention Block	Box	BP-mIOU	Box	BP-mIOU	CM-mIOU	SM-mIOU	PM-mIOU
✓	✓						75.1	64.9	76.2	66.6	-	-	-
✓		✓					-	-	76.2	-	48.1	-	-
✓	✓	✓					75.1	64.9	75.7	65.5	47.4	-	-
✓	✓	✓	✓				75.1	64.9	75.4	65.17	46.95	66.9	-
✓	✓	✓	✓	✓			75.1	64.9	75.15	65.01	46.56	66.42	69.7
✓	✓	✓	✓	✓	✓		75.5	64.97	76.39	66.76	48.65	66.89	70.08
✓	✓	✓	✓	✓	✓	✓	75.9	65.4	76.8	67.3	49.75	67.2	70.2

TABLE 7. Ablation studies were conducted on AF-CPACNet using various parsing instances and augmentations [27] including Body Parts parsing (BPM), Color Parsing (CP), Size Parsing (SP), and Pattern Parsing (PP) on CIHP and CCIHP datasets.

Architecutre	CIHP		CCIHP	
	Box AP0.5	BoxAP0.75	Box AP0.5	BoxAP0.75
RPRCNN [10]	93.44	80.12	-	-
ParsingRCNN [31]	93.68	80.25	-	-
AIParsing [34]	90.64	77.04	-	-
Ours	94.98	81.52	95.1	81.86

TABLE 8. Ablation Studies of Box Metrics for One-Stage Top-Down Architectures on CIHP and CCIHP Datasets.

CPACNet. Auto Augment is introduced to provide a new context and perspective to the network models for learning.

Ablation with Attention Network An increase of 2.25% in human part parsing accuracy and a 1.15% improvement in the model's ability to map different characteristics were observed after introducing augmentations and an additional attention network.

The AF-CPACNet demonstrates superior mIOU metric scores at all levels, highlighting its robustness, flexibility, and enhanced parsing and detection capabilities.

Limitations of the AF-CPACNet Approach
When multiple people are present in an image with very close proximity or in a highly crowded scene, there is a higher likelihood of mask bleeding occurring. This happens because the receptive fields of the feature maps may overlap, causing the model to confuse the boundaries between individuals. As a result, the predicted masks may bleed into adjacent regions, leading to inaccurate segmentation. This issue arises when the model struggles to differentiate between closely spaced objects due to the limited spatial resolution of the feature maps as shown in Figure 9.

VII. CONCLUSION & FUTURE WORKS

AF-CPACNet overcomes the limitations of traditional multi-stage anchor-based human parsing models by eliminating the need for anchor boxes and adopting a multi-task architecture. The model's use of a detection head and an edge-guided parsing module enables pixel-level analysis, which enhances the precision of human parsing. Additionally, the integration of a refinement head and a specialized loss function improves performance, particularly in capturing fine details such as color, size, and patterns at both bounding box and part levels. Experimental results demonstrate that AF-CPACNet significantly outperforms current state-of-the-art methods, with a notable 11% increase in mIOU on the CIHP dataset and achieving 75.9% mAP and 65.4% mIOU on the CCIHP dataset. This underscores AF-CPACNet's effectiveness in advancing human parsing, providing a robust solution for detailed and accurate analysis of human body parts. The model's design also allows for training and evaluation using single-stage networks, addressing production-level trade-off challenges and demonstrating versatility in practical applications. Future work could involve expanding AF-CPACNet's capabilities to handle a broader range of human poses and diverse clothing styles, as well as optimizing the model's efficiency and scalability for real-time scenarios and resource-constrained environments.

ACKNOWLEDGMENT

REFERENCES

[1] J Du, W Fu, Y Zhang, and Z Wang. Advancements in image recognition: A siamese network approach. Information Dynamics and Applications, 3(2):89–103, 2024.
[2] Riza Guler, Natalia Neverova, and Iasonas Kokkinos. Densepose: Dense human pose estimation in the wild. pages 7297–7306, 06 2018.
[3] Hussam J. Mohammed, Shumoos Al-Fahdawi, Alaa S. Al-Waisy, Dilovan Asaad Zebari, Dheyaa Ahmed Ibrahim, Mazin Abed Mohammed, Seifedine Kadry, and Jungeun Kim. Reid-deepnet: A hybrid deep learning system for person re-identification. Mathematics, 10(19), 2022.
[4] Z Xue and T Yao. Enhancing occluded pedestrian re-identification with the motionblur data augmentation module. Mechatron. Intell Transp. Syst, 3(2):73–84, 2024.
[5] Sonja Gievska and Sandra Treneska. Object detection and semantic segmentation of fashion images. 2020.
[6] SP Deore, TS Bagwan, PS Bhukan, HT Rajpal, and SB Gade. Enhancing image captioning and auto-tagging through a fcln with faster r-cnn integration. Inf. Dyn. Appl, 3:12–20, 2023.
[7] Zhaowei Cai and Nuno Vasconcelos. Cascade R-CNN: high quality object detection and instance segmentation. IEEE Trans. Pattern Anal. Mach. Intell., 43(5):1483–1498, 2021.
[8] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross B. Girshick. Pointrend: Image segmentation as rendering. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020, pages 9796–9805. Computer Vision Foundation / IEEE, 2020.
[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 2980–2988, 2017.
[10] Lu Yang, Qing Song, Zhihui Wang, Mengjie Hu, Chun Liu, Xueshi Xin, Wenhe Jia, and Songcen Xu. Renovating parsing r-cnn for accurate multiple human parsing. In European Conference on computer vision, pages 421–437. Springer, 2020.
[11] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. FCOS: fully convolutional one-stage object detection. CoRR, abs/1904.01355, 2019.
[12] Dangwei Li, Xiaotang Chen, Zhang Zhang, and Kaiqi Huang. Learning deep context-aware features over body and latent parts for person re-identification. 10 2017.
[13] Yu Chen, Ying Li, and Gang Wang. An enhanced region proposal network for object detection using deep learning method. PLOS ONE, 13:e0203897, 09 2018.
[14] Mahdi M. Kalayeh, Emrah Basaran, Muhittin Gokmen, Mustafa E. Kamasak, and Mubarak Shah. Human semantic parsing for person re-identification. CoRR, abs/1804.00216, 2018.
[15] Du Jiang, Gongfa Li, Ying Sun, Jiabing Hu, Juntong Yun, and Ying Liu. Manipulator grabbing position detection with information fusion of color image and depth image using deep learning. Journal of Ambient Intelligence and Humanized Computing, 12:10809–10822, 2021.
[16] Jian Zhao, Jianshu Li, Yu Cheng, Li Zhou, Terence Sim, Shuicheng Yan, and Jiashi Feng. Understanding humans in crowded scenes: Deep nested adversarial learning and a new benchmark for multi-human parsing. 07 2018.
[17] Sanyi Zhang, Xiaochun Cao, Guo-Jun Qi, Zhanjie Song, and Jie Zhou. AIParsing: Anchor-free instance-level human parsing. IEEE Transactions on Image Processing, 31:5599–5612, 2022.
[18] Dongli Wang, Shengliang Xiang, Yan Zhou, Jinzhen Mu, Haibin Zhou, and Richard Irampaye. Multiple-attention mechanism network for semantic segmentation. Sensors, 22(12), 2022.
[19] Zhen Zhou, Yan Zhou, Dongli Wang, Jinzhen Mu, and Haibin Zhou. Self-attention feature fusion network for semantic segmentation. Neurocomputing, 453:50–59, 2021.
[20] Michael Crawshaw. Multi-task learning with deep neural networks: A survey. CoRR, abs/2009.09796, 2020.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3501394

[21] Wei Peng, Qi Tang, Wei Dai, and Tielin Chen. Improving cancer driver gene identification using multi-task learning on graph convolutional network. Briefings in Bioinformatics, 23(1):bbab432, 2022.

[22] H El Bourakkadi, A Chemlal, H Tabti, M Kattass, A Jarjar, and A Benazzi. Enhanced color image encryption utilizing a novel vigenere method with pseudorandom affine functions. Acadlore Trans. Mach. Learn, 3(1):36–56, 2024.

[23] Ayah Hesham Abo El Rejal et al. An end-to-end cnn approach for enhancing underwater images using spatial and frequency domain techniques. 2023.

[24] YL Huang, N Li, and ZB Liu. An enhanced convolutional neural network for accurate classification of grape leaf diseases. Inf. Dyn. Appl, 2(1):8–18, 2023.

[25] I Hussain, J Muhammad, and R Ali. Enhanced global image segmentation: Addressing pixel inhomogeneity and noise with average convolution and entropy-based local factor. Int J. Knowl. Innov Stud, 1(2):116–126, 2023.

[26] S. K. Abhilash, Venu Madhav Nookala, Adithya Babu, Karthik S, and Mithun VR. MHCP-RCNN: multi-human color parsing segmentation using multi-task network. In International Joint Conference on Neural Networks, IJCNN 2023, Gold Coast, Australia, June 18-23, 2023, pages 1–8. IEEE, 2023.

[27] SK Abhilash, DS Nischal, Shiv Kumar, Venu Nookala, and S Karthik. Dcsp-rcnn: A network for person detection with color, size and pattern characteristic parsing. In 2023 International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS), pages 135–140. IEEE, 2023.

[28] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV), pages 801–818, 2018.

[29] Xiaomei Zhang, Yingying Chen, Bingke Zhu, Jinqiao Wang, and Ming Tang. Part-aware context network for human parsing. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 8971–8980, 2020.

[30] Peike Li, Yunqiu Xu, Yunchao Wei, and Yi Yang. Self-correction for human parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6):3260–3271, 2020.

[31] Lu Yang, Qing Song, Zhihui Wang, and Ming Jiang. Parsing r-cnn for instance-level human analysis. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 364–373, 2019.

[32] Lu Yang, Qing Song, Zhihui Wang, Zhiwei Liu, Songcen Xu, and Zhihao Li. Quality-aware network for human parsing. IEEE Transactions on Multimedia, 25:7128–7138, 2022.

[33] Angelique Loesch and Romaric Audigier. Describe me if you can! characterized instance-level human parsing. In 2021 IEEE International Conference on Image Processing (ICIP), pages 2528–2532. IEEE, 2021.

[34] Sanyi Zhang, Xiaochun Cao, Guo-Jun Qi, Zhanjie Song, and Jie Zhou. Aiparsing: Anchor-free instance-level human parsing. IEEE Transactions on Image Processing, 31:5599–5612, 2022.

[35] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. CoRR, abs/1606.00915, 2016.

[36] Jian Huang, Yingmei Wei, and Xiaolei Du. Pcnet: Exploratory visual analysis of large-scale network. In 2016 IEEE International Conference on Big Data Analysis (ICBDA), pages 1–6, 2016.

[37] Peike Li, Yunqiu Xu, Yunchao Wei, and Yi Yang. Self-correction for human parsing. CoRR, abs/1910.09777, 2019.

S. RAGHAVENDRA (Senior Member of IEEE), received a bachelor's degree in computer science and engineering from the BMS Institute of Technology, Bengaluru, a master's degree from the R. V. College of Engineering, Bengaluru, and a Ph.D. degree from the University Visvesvaraya College of Engineering, Bengaluru. He is currently working as an Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. He has 11 years of teaching and research experience in various institutes. He has authored more than 40 research publications. His research interests include cloud computing, machine learning, and the Internet of Things. He is currently serving as an Editorial Board Member, a Reviewer, and a Guest Editor for a number of prestigious journals of publishers, such as IEEE, Elsevier, Springer, Wiley, Taylor & Frances, and KJIP. He is serving as the Publication Chair for Discover-21. He was an Organizing Committee Member for conferences such as ICCN-14, ICCN 15, ICCN-16, ICInPro-18, DISCOVER-19, ICInPro-2019, and Discover-20. He also served as a Joint Secretary of the IEEE Mangalore Sub-Section, in 2020, and worked as the Website Co-Chair of IEEE MSS, in 2019.

S. K. ABHILASH has received the Master of Engineering in electronics and communication engineering from UVCE, Bangalore. He is currently working as an Senior Technical Lead at PathPartner member of KPIT Group . He is working on developing AI architecture models for different computer vision-based use cases for Automotive, Surveillance and Medical Imaging domains. He has published around 10 papers in international journals and conferences along with a patent to his credits. His research work focuses on Advanced Driver Assistance Systems (ADAS), Surveillance analytics systems, Diffusion models, Generative AI, Annotation Tools and building unified agnostic-based computer vision frameworks.

VENU MADHAV NOOKALA has graduated from Amrita School of Engineering, Coimbatore. He has publications more than five papers in the top conferences. He is working as a Software Engineer in PathPartner member of KPIT Group, Bangalore. He has started his career on Computer vision algorithms, is an enthusiastic artificial intelligence engineer seeking to deliver state-of-the-art AI solutions capable of learning their own structures to analyze events.His current work focuses on AI based Automotive and surveillance analytics system, developing AI based tools based on graphic user interface, Annotation Tools, Deployment of Computer vision models on various frameworks and developing the multi-task architectures.

YASHWANTH NANJAPPA (Senior Member of IEEE), received the B.E. degree in Electronics and Communication Engineering from Visvesvaraya Technological University, Belagavi, in 2010, the M.Tech. degree in Signal Processing and VLSI from Jain University, Bengaluru, in 2012, and the Ph.D. degree from Visvesvaraya Technological University in 2020. He is currently an Assistant Director (FDW) and an Assistant Professor (Senior Scale) with the Department of Electronics and Communication Engineering at the Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, where he is actively involved in numerous administrative and academic roles. His primary research interests include wireless sensor networks, communication systems, antennas, VLSI technology, and cyber security.