applied sciences

MDPI

*Article*

# Chromosome Cluster Type Identification Using a Swin Transformer

**Indu Joshi [1,†], Arnab Kumar Mondal [2,†] and Nassir Navab [1,*]**

[1] Computer Aided Medical Procedures (CAMP), Technical University of Munich, 85748 Munich, Germany; indu.joshi@tum.de
[2] Indian Institute of Technology Delhi, New Delhi 110016, India; anz188380@iitd.ac.in
* Correspondence: navab@cs.tum.edu
† These authors contributed equally to this work.

**Abstract:** The analysis of chromosome karyotypes is crucial for diagnosing genetic disorders such as Patau syndrome, Edward syndrome, and Down syndrome. Chromosome cluster type identification is a key step in the automated analysis of chromosome karyotypes. State-of-the-art chromosome cluster-type identification techniques are based on convolutional neural networks (CNNs) and fail to exploit the global context. To address this limitation of the state of the art, this paper proposes a transformer network, *c*hromosome *c*luster *t*ransformer (CCT), that exploits a swin transformer backbone and successfully captures long-range dependencies in a chromosome image. Additionally, we find that the proposed CCT has a large number of model parameters, which makes it prone to overfitting on a (small) dataset of chromosome images. To alleviate the limited availability of training data, the proposed CCT also utilizes a transfer learning approach. Experiments demonstrate that the proposed CCT outperforms the state-of-the-art chromosome cluster type identification methods as well as the traditional vision transformer. Furthermore, to provide insights on the improved performance, we demonstrate the activation maps obtained using Gradient Attention Rollout.

**Keywords:** chromosome cluster identification; chromosome karyotype analysis; deep learning; transformer; transfer learning



**Citation:** Joshi, I.; Mondal, A.K.; Navab, N. Chromosome Cluster Type Identification Using a Swin Transformer. *Appl. Sci.* **2023**, *13*, 8007. https://doi.org/10.3390/app13148007

Academic Editors: Andrea Ballini and Vincent A. Cicirello

Received: 28 March 2023
Revised: 27 June 2023
Accepted: 4 July 2023
Published: 8 July 2023



## 1. Introduction

Chromosomes contain the genetic information in a human cell [1]. Chromosome karyotype analysis facilitates the prenatal diagnosis of severe abnormalities or genetic disorders such as Patau syndrome, Edward syndrome, and Down syndrome [2,3]. Given a grayscale image of a stained cell, the process of karyotype analysis is characterized by the segmentation of chromosome instances and the subsequent arrangement of the karyotypes according to their respective categories. Karyotype analysis is generally carried out by experienced medical practitioners. However, the manual analysis of karyotypes is tedious and highly time-consuming [4–6]. Furthermore, there can be inconsistencies in the process due to the required domain expertise or even due to the fatigue and over-burden of the medical practitioners. The above-mentioned limitations of the manual analysis of chromosome karyotype analysis motivate the need to design an automated decision support system for chromosome karyotype analysis [7–9]. An automated karyotype analysis model has two stages: *chromosome segmentation* and *chromosome classification*. As shown in Figure 1, chromosome cluster type identification is a key step towards chromosome segmentation and classification [10,11].

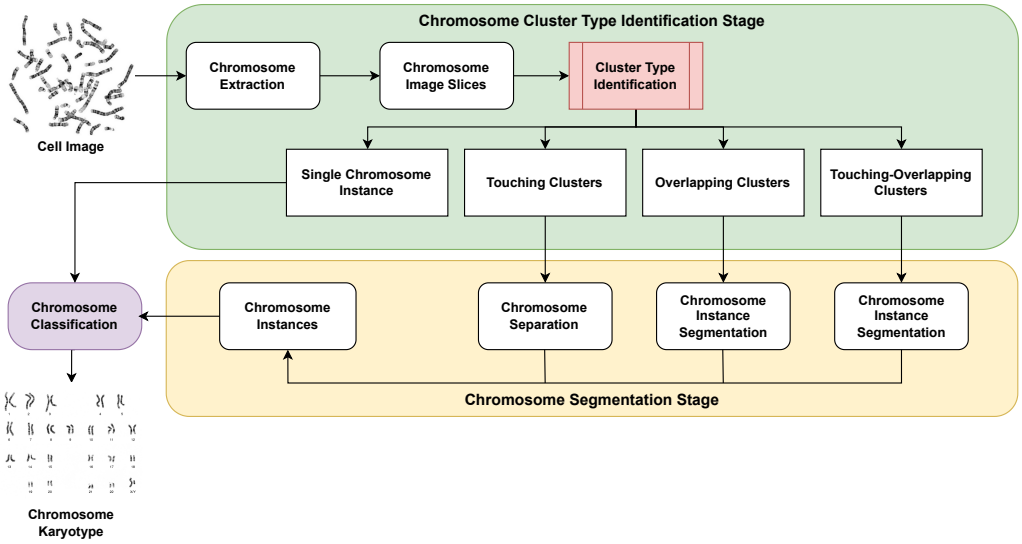



**Figure 1.** Flowchart depicting the pipeline of an automated karyotype analysis model. The model consists of *chromosome segmentation* and *chromosome classification* stage. The contribution of this paper lies in proposing a *cluster type identification* module (marked in red) that helps to perform the required chromosome segmentation step.

Recent methods for chromosome cluster type identification [10,11] exploit deep models based on convolutional neural networks (CNNs). However, CNNs are known to exhibit inductive bias and fail to capture long-range dependencies [12]. Recently, transformers have been introduced to circumvent the aforementioned shortcomings of CNNs [12]. Motivated by the success of transformers in image processing applications [13–16], we introduce chromosome cluster transformer (CCT)—a swin transformer [17]-based hierarchical transformer model that leverages self-attention to identify cluster type in chromosome images. To summarize, this paper makes the following research contributions:

- We address the limitation of state-of-the-art CNN-based chromosome cluster type identification methods in capturing long-range dependencies. Towards this end, we propose the *c*hromosome *c*luster *t*ransformer (CCT), which successfully captures the global context required for the successful identification of chromosome cluster types.
- To the best of our knowledge, this is the first work in the domain of chromosome cluster type identification that utilizes a transformer model.
- To circumvent the limited availability of training data for cluster type identification, the proposed CCT exploits a transfer learning approach.
- The proposed CCT outperforms the state-of-the-art traditional vision transformer in chromosome cluster type identification.
- Furthermore, the proposed CCT outperforms the existing state-of-the-art chromosome cluster type identification methods.
- Additionally, to provide insights on the improved performance, we visualize the activation maps obtained using Gradient Attention Rollout [18].

## 2. Related Work

The automated analysis of chromosome images has intrigued researchers for a long time [19–24]. The earliest methods for chromosome cluster identification exploited geometric features [25–27]. Minaee et al. [25] proposed a geometric method that exploits cut-line to segment touching and partially overlapping chromosomes. Kubola and Wayalun [26] utilized geometric features such as intersection points, as well as endpoints corresponding to the image skeleton of the given chromosome image obtained after preprocessing. Arora and Dhir [27] exploited the geometric features of an object, such as its circularity, area, and length. Recently, learning-based methods have been exploited for the analysis

of chromosome images [8,10,11,22,28–30]. Lin et al. [11] compared the performance of classical machine learning and deep learning methods on the chromosome cluster type identification task. One of the most recent methods proposed for automated cluster type identification in chromosome images is the ResNeXt-WSL [10] model.

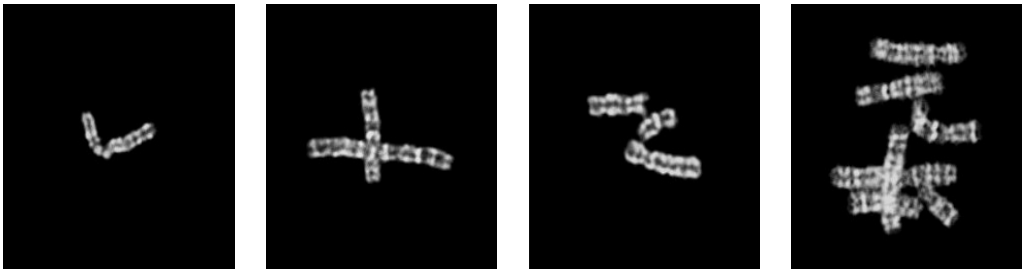
A promising direction for the segmentation of chromosome instances is the use of *co-saliency detection*. The process of locating and highlighting similar salient regions in an image is known as co-saliency detection. It is a useful technique to identify a group of related image regions [31]. The different research directions explored for co-saliency detection can be broadly categorized as optimization-based, graph-based, and deep-learning-based techniques. When posed as an optimization problem, sparse coding [32] and matrix factorization [33] have been used for co-saliency detection. Graph-based co-saliency detection techniques exploit graphs to represent the relationship between different image regions and identify the related image regions [34,35]. On the other hand, deep-learning-based techniques learn to predict the co-saliency maps by employing deep neural networks [36,37]. For this research, we keep our focus limited to the classification of chromosome images and overcome the inability of state-of-the-art [10] approaches to understand long-range dependencies and the global context by introducing a transformer-based classification model.

We hypothesize that due to high inter-class similarity, global context is crucial to correctly identify cluster types in the chromosome images as some classes, for instance, touching chromosome cluster and touching–overlapping chromosome cluster images. However, we observe that all the deep models studied so far for chromosome cluster identification are CNN-based models, and so suffer from the inductive bias observed in CNNs and, more importantly, fail to capture long-range dependencies and global context. We address this limitation of state-of-the-art cluster-type identification methods by proposing a self-attention-based hierarchical model that captures the global context in a chromosome image and thereby allows better performance compared to the state of the art.

## 3. Proposed Method

### *3.1. Problem Formulation*

Depending upon the connectivity of chromosome instances, a given chromosome image denoted as $X$ is labeled as one of four classes: chromosome instance, overlapping chromosome cluster, touching chromosome cluster, or touching–overlapping chromosome cluster. Thus, cluster type identification is formulated as a four-class classification problem. Samples of all four class types are provided in Figure 2.



(**a**) (**b**) (**c**) (**d**)

**Figure 2.** Samples from [10] illustrating all four categories of cluster types: (**a**) single chromosome instance, (**b**) overlapping chromosome cluster, (**c**) touching chromosome cluster, (**d**) touching–overlapping chromosome cluster.

### *3.2. Chromosome Cluster Transformer (CCT)*

The proposed CCT is a self-attention guided deep model based on a swin transformer [17] that captures the global context of a chromosome image by modeling long-range dependencies. However, different from the traditional vision image transformer [12], the proposed CCT is a hierarchical model that computes multi-scale representations. At the first stage of the proposed CCT, from a given chromosome image $x \in R^{H \times W \times C}$, $N$ patches

of size $4 \times 4$ are extracted, where $N = \frac{H \times W}{4 \times 4}$. The extracted patches are projected onto a linear layer to provide a C-dimensional embedding for each patch. In the second stage, in order to obtain a hierarchical representation, patches in the $2 \times 2$ neighborhoods are merged to obtain a 4C dimensional feature representation, which is projected onto the linear layer to obtain a total of $\frac{H \times W}{8 \times 8}$ patches, each with a feature vector of length 2C. A similar procedure of patch merging and feature transformation is followed at the third and the fourth stage to obtain a total of 4C dimensional $\frac{H \times W}{16 \times 16}$ patches and 8C dimensional $\frac{H \times W}{32 \times 32}$ patches, respectively. All four stages jointly provide a hierarchical representation (see Figure 3a) that is used to identify cluster types in a given chromosome image.

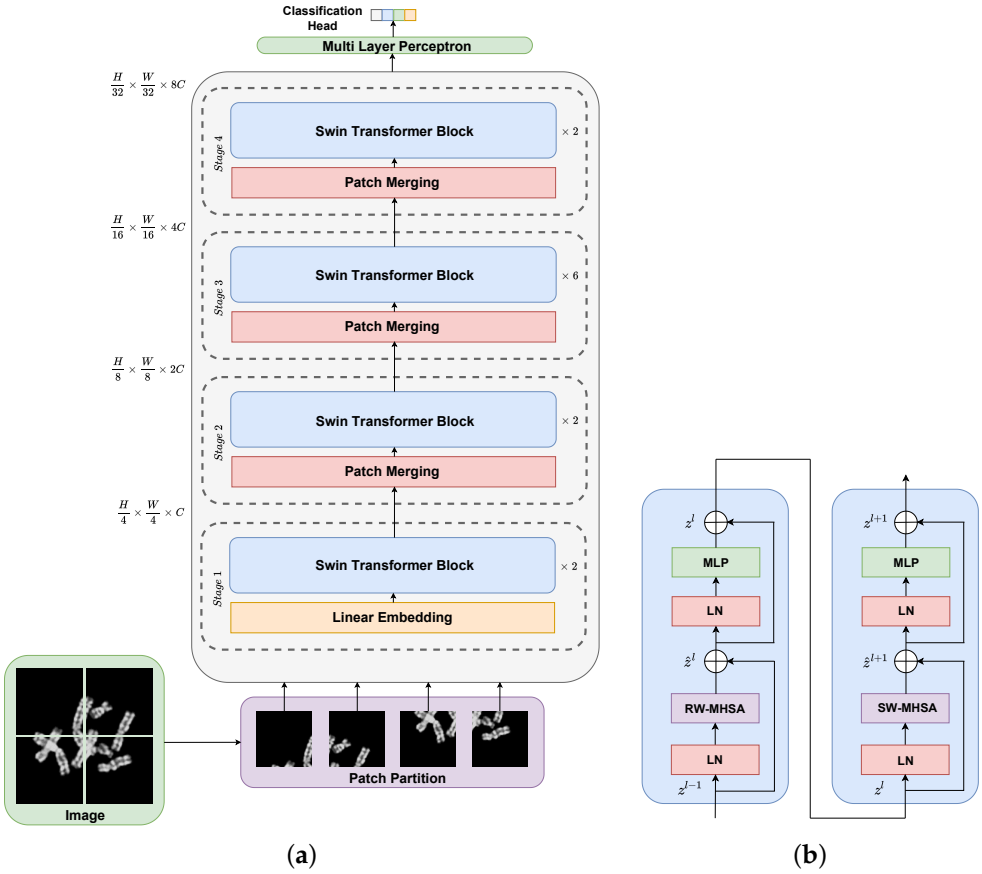



**Figure 3.** Flowchart depicting (**a**) a schematic diagram of the proposed chromosome cluster transformer (CCT), and (**b**) the architecture of a constituting swin transformer block. MLP, multi-layer perceptron; LN, layernorm layer; RW-MHSA, regular window-based multi-head self-attention; and SW-MHSA, shifted window-based multi-head self-attention.

The traditional vision transformer exploits a window-based self-attention mechanism that does not interact among the different non-overlapping windows, leading to a limited representation ability of the model. To circumvent this limitation, the proposed CCT exploits two self-attention modules: a traditionally used regular window-based multi-head self-attention head (RW-MHSA) and a shifted window-based multi-head self-attention (SW-MHSA) head that effectively introduces cross-window connections (see Figure 3b).

### *3.3. Partitioning of Shifted Windows*

The traditional vision transformer exploits a window-based multi-head self-attention module (MHSA). However, this standard MHSA module does not have any connections across the non-overlapping windows. To increase the model capacity, the proposed CCT exploits cross-window connections through the partitioning of shifted windows. Through the shifted window partitioning mechanism, two different partitioning configurations are

maintained between the two adjacent swin transformer blocks [17]. As a result, the first attention module exploits regular window partitioning such that the image is partitioned from the pixel at the top left, and the corresponding $8 \times 8$ feature map is evenly partitioned into $4 \times 4$ sized $2 \times 2$ windows. However, the subsequent attention module exploits the shifted window partitioning such that the window configuration in this module is obtained by displacing the windows of the previous layer by (2,2) pixels. A two-layer multi-layer perceptron (MLP) follows the transformer blocks in the proposed CCT. Every MHSA module and MLP have a LayerNorm (LN) layer applied prior to them, and every module also has a residual connection added after it. Computations by the proposed CCT can be formally defined as

$$\hat{z}^l = \text{RW-MHSA}(LN(z^{l-1})) + z^{l-1} \quad (1)$$

$$z^l = \text{MLP}(LN(\hat{z}^l)) + \hat{z}^l \quad (2)$$

$$\hat{z}^{l+1} = \text{SW-MHSA}(LN(z^l)) + z^l \quad (3)$$

$$z^{l+1} = \text{MLP}(LN(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (4)$$

### *3.4. Transfer Learning for Chromosome Cluster Transformer*

Transformers lack inductive bias, such as locality and translation equivariance, due to which transformers require larger training datasets compared to CNNs. To alleviate the limited availability of training data for training the proposed CCT, we exploit a transfer learning approach. We initialize the proposed CCT with the pre-trained weights of the swin transformer trained on the large-scale ImageNet database [38]. Subsequently, the proposed CCT is fine-tuned on the chromosome image training dataset.

### *3.5. Implementation Details*

The proposed CCT is implemented on Tensorflow 2.x. The network is trained using the Adam optimizer over a cross-entropy loss with label smoothing. A batch size of 16, a patch size of 16, and a learning rate of 0.0001 are used. The model features a GPU node with one Nvidia V100 card and two Intel Xeon G-6148 CPUs. While preprocessing, the images are resized to $384 \times 384$, and the augmentations used include rotation, width shift, height shift, shear, and zoom. Additional augmentations used include horizontal flip, vertical flip, and varying the image brightness.

## 4. Databases and Experimental Protocol

### *4.1. Database*

The proposed CCT is evaluated on a publicly available clinical database [10]. The database was collected from the Medical Genetic Centre and Maternal and Children Metabolic–Genetic Key Laboratory of Guangdong Women and Children Hospital [10]. The dataset comprises 500 stained microphotograph cell images that were eventually segmented into 6592 chromosome images by the authors. The samples were manually classified by the authors into one of four classes: chromosome instance, overlapping cluster, touching cluster, and touching–overlapping cluster. The distribution of the class labels in the database is summarized in Table 1. Furthermore, as the dataset is not explicitly divided into training and testing sets, similar to the authors of [10], we exploit the hold-out technique to determine the model performance.

**Table 1.** Distribution of class labels in the database [10].

| Class Label | Image Count |
|---|---|
| Chromosome Instance | 1712 |
| Overlapping Cluster | 1038 |
| Touching Cluster | 3029 |
| Touching–Overlapping Cluster | 813 |

### 4.2. Evaluation Metrics

To assess the classification performance obtained by the proposed CCT, the following evaluation metrics were exploited. Let $TN$ and $TP$ denote the total number of true negative and true positive samples classified by the proposed CCT. Similarly, $FN$ and $FP$ denote the total number of false negative and false positive samples classified by the proposed CCT. $N$ denotes the total number of image samples in the database. The evaluation metrics used in this paper are formally defined as follows:

1. *Accuracy:*

$$Accuracy = \frac{(TP + TN)}{N} \tag{5}$$

2. *Precision:*

$$Precision = \frac{TP}{(TP + FP)} \tag{6}$$

3. *F1 score:*

$$F1score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{7}$$

4. *Sensitivity:*

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{8}$$

5. *Specificity:*

$$Specificity = \frac{TN}{(TN + FP)} \tag{9}$$

## 5. Results and Analysis

### 5.1. Comparison with State of the Art

We begin the analysis of the proposed CCT by comparing its performance with seven state-of-the-art models for cluster type identification. A 5-fold cross-validation is performed only for the proposed CCT and the ResNeXt-WSL [10], and the average value of the results obtained for different folds is taken to represent the results of the corresponding method. For the rest of the baselines, due to the limited availability of computational resources, the numbers reported in [10,39] are taken. For all the baselines, pre-trained models trained on ImageNet dataset are used. The corresponding classification results are reported in Table 2. The proposed CCT significantly outperforms all seven state-of-the-art cluster-type identification models on all five evaluation metrics, demonstrating the improved performance obtained by the proposed CCT.

**Table 2.** Comparison of the proposed CCT with the state of the art. Bold values represent the best results.

| Model | Precision | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|---|
| MobileNetV2 [40] | 81.83 | 83.41 | 77.52 | 76.85 | 94.47 |
| DenseNet121 [41] | 85.59 | 87.65 | 82.23 | 81.68 | 95.88 |
| ResNet-50 [42] | 88.30 | 90.15 | 86.08 | 85.68 | 96.72 |
| ResNet-101 [42] | 90.65 | 91.89 | 88.32 | 87.92 | 97.30 |
| ResNet-152 [42] | 90.71 | 91.97 | 89.09 | 88.79 | 97.32 |
| ResNeXt-101-32×8d [43] | 90.79 | 92.27 | 89.36 | 89.10 | 97.42 |
| ResNeXt-WSL [10] | 93.35 | 94.13 | 92.41 | 92.20 | 98.04 |
| Dual-ViT [44] | 94.07 | 94.10 | 94.05 | 94.10 | 97.69 |
| SupCAM [39] | 93.25 | 94.99 | 92.26 | 92.81 | 98.12 |
| **CCT (Proposed)** | **95.02** | **95.30** | **95.02** | **95.03** | **98.26** |

### *5.2. Cross-Validation Performance*

Next, we assess the stability and robustness of the proposed CCT for different test samples by performing five-fold cross-validation. To perform five-fold cross-validation, the dataset is randomly split into five folds. Iteratively, one fold is selected as the validation set, while the samples corresponding to the rest of the four folds are selected as the training set. The model's performance is evaluated for each of the five validation sets and reported in Table 3. We observe that the proposed CCT performs well for the different choices of training and validation sets, demonstrating the robustness and reliability of the proposed CCT on different choices of training and validation data.

**Table 3.** Cross-validation performance of the proposed CCT compared to ResNeXt-WSL [10].

| Method | Fold | Precision | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| ResNeXt-WSL [10] | 1 | 94.28 | 94.77 | 93.58 | 93.43 | 98.26 |
| | 2 | 91.52 | 92.73 | 90.54 | 90.33 | 97.58 |
| | 3 | 92.70 | 93.63 | 91.21 | 90.82 | 97.88 |
| | 4 | 94.42 | 94.99 | 93.42 | 93.23 | 98.33 |
| | 5 | 93.82 | 94.53 | 93.31 | 93.20 | 98.18 |
| | Mean (±std) | 93.35 ± 2.19 | 94.13 ± 1.68 | 92.41 ± 2.55 | 92.20 ± 2.68 | 98.04 ± 0.56 |
| **CCT (Proposed)** | 1 | 95.37 | 95.75 | 95.36 | 95.36 | 98.44 |
| | 2 | 94.52 | 94.84 | 94.55 | 94.59 | 98.09 |
| | 3 | 95.00 | 95.07 | 95.02 | 95.01 | 98.15 |
| | 4 | 94.84 | 95.29 | 94.83 | 94.81 | 98.29 |
| | 5 | 95.37 | 95.53 | 95.36 | 95.36 | 98.32 |
| | **Mean (±std)** | **95.02 ± 0.32** | **95.30 ± 0.31** | **95.02 ± 0.31** | **95.03 ± 0.30** | **98.26 ± 0.12** |

### *5.3. Confusion Matrix*

The results reported so far indicate the overall classification performance for all the classes. Next, we assess the per-class classification performance. To achieve this, we plot the confusion matrices obtained for all five folds (see Figure 4). We observe that across all folds, on average, the touching class appears to be the most challenging for classification and is often misclassified as the overlapping or instance class. Similarly, we find that the overlapping class, when misclassified, is most likely to be classified as the instance class. These results are intuitive as such confusions across classes are likely to be made by a human expert as well.

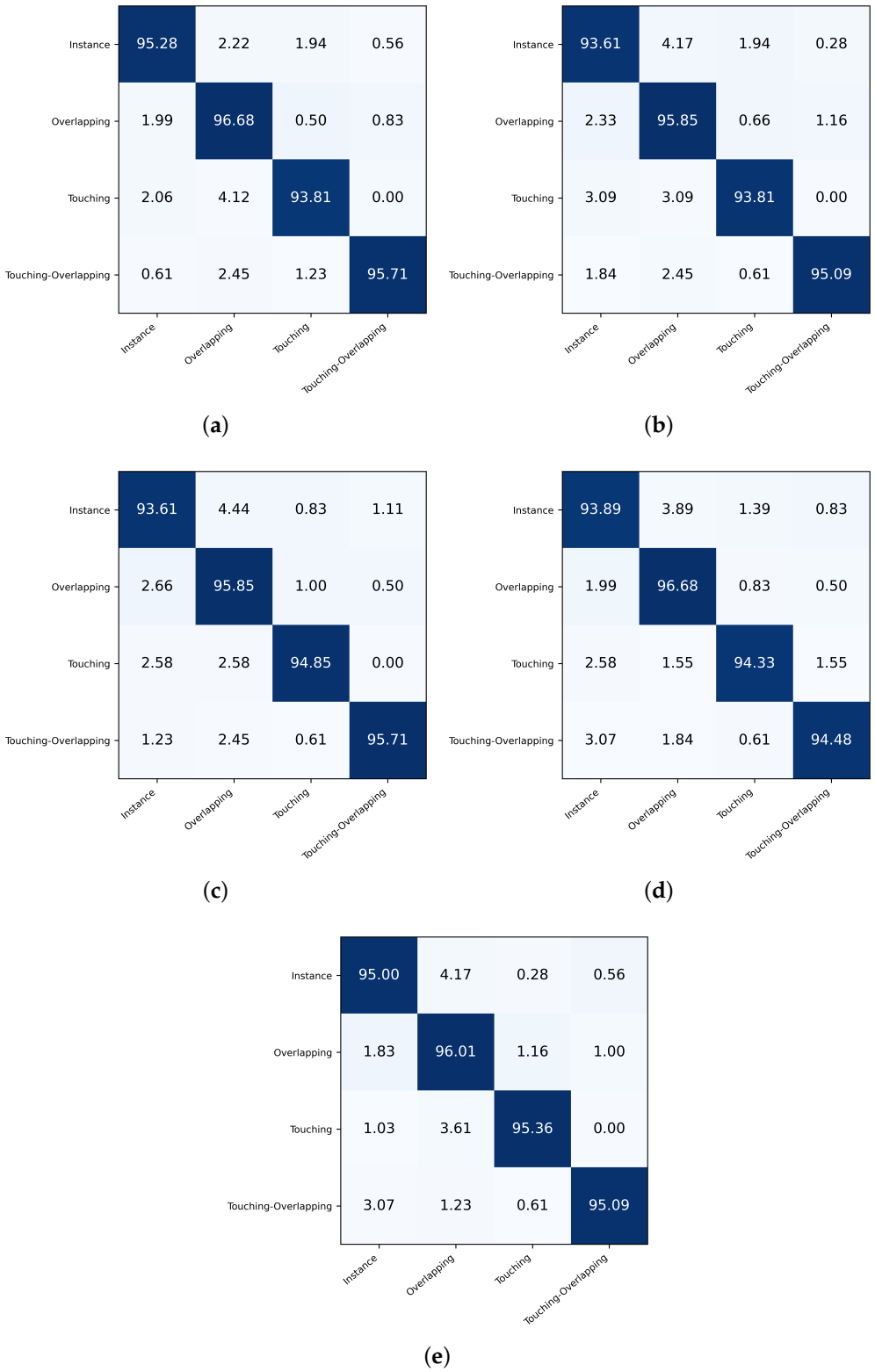



**Figure 4.** Confusion matrices quantifying the per class classification performance of the proposed CCT for each of the five folds curated during the cross-validation of the proposed CCT. Subfigures (**a**–**e**) represent the confusion matrices obtained respectively for the first to the fifth fold.

### *5.4. Effect of Model Architecture*

Next, we study the impact of the model architecture and the number of parameters on the performance of the proposed CCT. The default architecture is denoted as CCT-Large.

Different variants of the proposed CCT include CCT-Tiny, CCT-Small, and CCT-Base, which constitute approximately 0.125×, 0.250×, and 0.500× the model parameters compared to the default architecture (CCT-Large). The difference in CCT-Tiny, CCT-Small, CCT-Base, and CCT-Large exists in the number of stages and the number of transformer blocks in each stage. CCT-Tiny consists of 2 stages, each with four transformer blocks and a latent representation of dimension 96. CCT-Small constitutes two stages, each with four transformer blocks and a latent representation of dimension 96. CCT-Base comprises four stages, each with four transformer blocks and a latent representation of dimension 128. CCT-Large comprises four stages, each with four transformer blocks and a latent representation of dimension 192. Subsequently, the model with a greater number of stages and transformer encoder layers, and a higher dimensionality of latent representation of the transformer encoder has more parameters. While CCT-Base contains 88 M trainable parameters, CCT-Tiny, CCT-Small, and CCT-Large contain 29 M, 50 M, and 197 M parameters, respectively. As expected, the classification improves as the model capacity increases (see Table 4). We observe that CCT-Large achieves the best classification performance. Therefore, we adopt CCT-Large as the default architecture for all the experiments reported in this paper.

**Table 4.** Effect of choice of model architecture on the performance of the proposed CCT. Bold values represent the best results.

| Model | Precision | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|---|
| CCT-Tiny | 94.11 ± 0.31 | 94.72 ± 0.33 | 94.00 ± 0.29 | 94.05 ± 0.27 | 97.99 ± 0.34 |
| CCT-Small | 94.61 ± 0.20 | 95.16 ± 0.21 | 94.51 ± 0.25 | 94.55 ± 0.19 | 98.29 ± 0.23 |
| CCT-Base | 94.99 ± 0.15 | 95.13 ± 0.81 | 94.89 ± 0.11 | 95.01 ± 0.36 | 98.12 ± 0.12 |
| CCT-Large | **95.02 ± 0.32** | **95.30 ± 0.31** | **95.02 ± 0.31** | **95.03 ± 0.30** | **98.26 ± 0.12** |

### *5.5. Comparison with Traditional Vision Transformer*

Next, we compare the classification performance of the proposed CCT with a traditional vision transformer [12]. Contrary to ViT, the proposed CCT employs a hierarchical architecture to accommodate multi-scale information and capture long-range dependencies. As a result, as reported in Table 5, the proposed CCT significantly outperforms the ViT model.

**Table 5.** Comparison of the proposed CCT with the traditional vision transformer model. Bold values represent the best results.

| Model | Precision | Accuracy | F1 | Sensitivity | Specificity |
|---|---|---|---|---|---|
| ViT [12] | 90.17 | 92.75 | 91.06 | 91.97 | 97.34 |
| **CCT (Proposed)** | **95.02** | **95.30** | **95.02** | **95.03** | **98.26** |

### *5.6. Visualization of Model Activation*

Lastly, to qualitatively analyze the improved performance obtained by the proposed CCT and obtain insights into the salient regions that help CCT to predict a class, we visualize its activation maps. The activation maps are computed using Gradient Attention Rollout [18], a state-of-the-art method to visualize the activation maps of a transformer model. Figure 5 demonstrates that higher activation is obtained for salient and decisive image regions, such as the touching region for the touching and touching–overlapping classes. The higher activation around informative and decisive image regions explains the superior classification performance obtained by the proposed CCT.

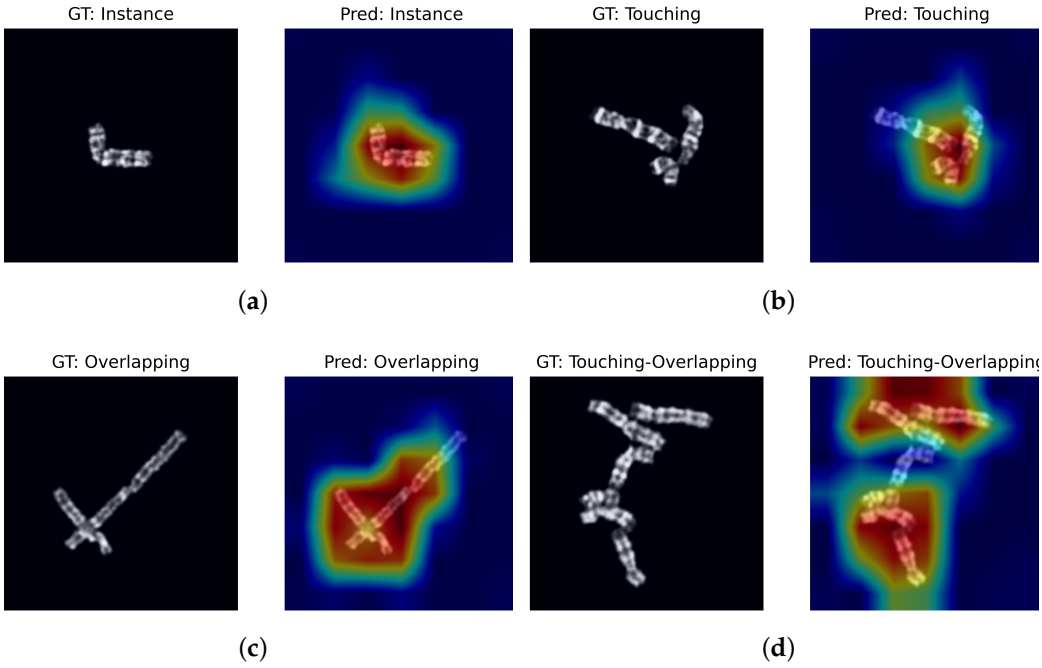



**Figure 5.** Visualization of activation maps of the proposed CCT. As expected, CCT obtains higher activation around salient and decisive image regions for all four classes. The higher activation around decisive image regions helps to qualitatively analyze the superior classification performance of the proposed CCT. Subfigures (**a**–**d**) demonstrate the respective activation maps obtained for the samples of the four classes.

## 6. Conclusions

This research introduces the *c*hromosome *c*luster *t*ransformer (CCT) to classify a given chromosome into one of four classes: instance, touching, overlapping, or touching–overlapping. The results demonstrate the superior classification performance of the proposed CCT compared to the state of the art. Furthermore, the visualization of model activation maps provides the finding that higher activation is obtained around decisive and more informative image regions, which subsequently helps the proposed CCT to obtain superior classification performance. Chromosome cluster type identification is the first step toward chromosome segmentation. Therefore, the proposed method can be viewed as the first step in the direction of achieving this goal. In the future, we intend to extend this idea and develop an end-to-end model that, given an input image, will simultaneously perform both cluster type identification and segmentation.

**Author Contributions:** Conceptualization, A.K.M. and I.J.; methodology, A.K.M. and I.J.; software, A.K.M. and I.J.; writing—original draft preparation, A.K.M. and I.J.; writing—review and editing, A.K.M., I.J. and N.N.; visualization, A.K.M. and I.J.; supervision, N.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** A. Mondal is supported by the Prime Minister's Research Fellows scheme of the government of India. This work was done as a part of the IMI BigPicture project (IMI945358).

**Institutional Review Board Statement:** Ethical review and approval were waived for this study as the study was conducted on a public available dataset.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

**Data Availability Statement:** Publicly available dataset was analyzed in this study. This data can be found here: [https://doi.org/10.1016/j.media.2020.101943].

**Acknowledgments:** The authors acknowledge support from the HPC facility of IIT Delhi.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References


1. Rai, A.; Hirakawa, H.; Rai, M.; Shimizu, Y.; Shirasawa, K.; Kikuchi, S.; Seki, H.; Yamazaki, M.; Toyoda, A.; Isobe, S.; et al. Chromosome-scale genome assembly of Glycyrrhiza uralensis revealed metabolic gene cluster centred specialized metabolites biosynthesis. *DNA Res.* **2022**, *29*, dsac043. [CrossRef] [PubMed]
2. Wang, Y.P.; Wu, Q.; Castleman, K.R.; Xiong, Z. Chromosome image enhancement using multiscale differential operators. *IEEE Trans. Med. Imaging* **2003**, *22*, 685–693. [CrossRef] [PubMed]
3. Qin, Y.; Wen, J.; Zheng, H.; Huang, X.; Yang, J.; Song, N.; Zhu, Y.M.; Wu, L.; Yang, G.Z. Varifocal-net: A chromosome classification approach using deep convolutional networks. *IEEE Trans. Med. Imaging* **2019**, *38*, 2569–2581. [CrossRef]
4. Arora, T.; Dhir, R. A review of metaphase chromosome image selection techniques for automatic karyotype generation. *Med. Biol. Eng. Comput.* **2016**, *54*, 1147–1157. [CrossRef] [PubMed]
5. Remani Sathyan, R.; Chandrasekhara Menon, G.; Thampi, R.; Duraisamy, J.H. Traditional and deep-based techniques for end-to-end automated karyotyping: A review. *Expert Syst.* **2022**, *39*, e12799. [CrossRef]
6. Remya, R.; Hariharan, S.; Keerthi, V.; Gopakumar, C. Preprocessing G-banded metaphase: Towards the design of automated karyotyping. *SN Appl. Sci.* **2019**, *1*, 1–8. [CrossRef]
7. Wei, H.; Gao, W.; Nie, H.; Sun, J.; Zhu, M. Classification of Giemsa staining chromosome using input-aware deep convolutional neural network with integrated uncertainty estimates. *Biomed. Signal Process. Control* **2022**, *71*, 103120. [CrossRef]
8. Huang, K.; Lin, C.; Huang, R.; Zhao, G.; Yin, A.; Chen, H.; Guo, L.; Shan, C.; Nie, R.; Li, S. A novel chromosome instance segmentation method based on geometry and deep learning. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8.
9. Menaka, D.; Vaidyanathan, S.G. Chromenet: A CNN architecture with comparison of optimizers for classification of human chromosome images. *Multidimens. Syst. Signal Process.* **2022**, *33*, 747–768. [CrossRef]
10. Lin, C.; Zhao, G.; Yin, A.; Yang, Z.; Guo, L.; Chen, H.; Zhao, L.; Li, S.; Luo, H.; Ma, Z. A novel chromosome cluster types identification method using ResNeXt WSL model. *Med. Image Anal.* **2021**, *69*, 101943. [CrossRef]
11. Lin, C.; Yin, A.; Wu, Q.; Chen, H.; Guo, L.; Zhao, G.; Fan, X.; Luo, H.; Tang, H. Chromosome cluster identification framework based on geometric features and machine learning algorithms. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 2357–2363.
12. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
13. Hu, R.; Chen, J.; Zhou, L. A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. *Comput. Biol. Med.* **2022**, *144*, 105325. [CrossRef]
14. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef] [PubMed]
15. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in vision: A survey. *ACM Comput. Surv. CSUR* **2021**, *54*, 1–41. [CrossRef]
16. Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; Landman, B.; Roth, H.R.; Xu, D. Unetr: Transformers for 3d medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2022; pp. 574–584.
17. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
18. Chefer, H.; Gur, S.; Wolf, L. Transformer interpretability beyond attention visualization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 782–791.
19. Kimori, Y. Morphological image processing for quantitative shape analysis of biomedical structures: effective contrast enhancement. *J. Synchrotron Radiat.* **2013**, *20*, 848–853. [CrossRef] [PubMed]
20. Ming, D.; Tian, J. Automatic pattern extraction and classification for chromosome images. *J. Infrared Millim. Terahertz Waves* **2010**, *31*, 866–877. [CrossRef]
21. Altinsoy, E.; Yang, J.; Yilmaz, C. Fully automatic raw G-band chromosome image segmentation. *IET Image Process.* **2020**, *14*, 1920–1928. [CrossRef]
22. Liu, X.; Fu, L.; Chun-Wei Lin, J.; Liu, S. SRAS-net: Low-resolution chromosome image classification based on deep learning. *IET Syst. Biol.* **2022**, *16*, 85–97. [CrossRef]
23. Arora, T.; Dhir, R. A variable region scalable fitting energy approach for human Metaspread chromosome image segmentation. *Multimed. Tools Appl.* **2019**, *78*, 9383–9404. [CrossRef]

24. Madian, N.; Jayanthi, K.; Suresh, S. Analysis of human chromosome images: Application towards an automated chromosome classification. *Int. J. Imaging Syst. Technol.* **2018**, *28*, 235–245. [CrossRef]
25. Minaee, S.; Fotouhi, M.; Khalaj, B.H. A geometric approach to fully automatic chromosome segmentation. In Proceedings of the 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 13 December 2014; pp. 1–6.
26. Kubola, K.; Wayalun, P. Automatic determination of the g-band chromosomes number based on geometric features. In Proceedings of the 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhonpathom, Thailand, 11–13 July 2018; pp. 1–5.
27. Arora, T.; Dhir, R. A novel approach for segmentation of human metaphase chromosome images using region based active contours. *Int. Arab. J. Inf. Technol.* **2019**, *16*, 132–137.
28. Sharma, M.; Vig, L. Automatic chromosome classification using deep attention based sequence learning of chromosome bands. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
29. Saleh, H.M.; Saad, N.H.; Isa, N.A.M. Overlapping chromosome segmentation using u-net: Convolutional networks with test time augmentation. *Procedia Comput. Sci.* **2019**, *159*, 524–533. [CrossRef]
30. Altinsoy, E.; Yang, J.; Tu, E. An improved denoising of G-banding chromosome images using cascaded CNN and binary classification network. *Vis. Comput.* **2021**, *38*, 2139–2152. [CrossRef]
31. Zhang, D.; Fu, H.; Han, J.; Borji, A.; Li, X. A review of co-saliency detection algorithms: Fundamentals, applications, and challenges. *ACM Trans. Intell. Syst. Technol. TIST* **2018**, *9*, 1–31. [CrossRef]
32. Huang, R.; Feng, W.; Sun, J. Color feature reinforcement for cosaliency detection without single saliency residuals. *IEEE Signal Process. Lett.* **2017**, *24*, 569–573. [CrossRef]
33. Li, T.; Song, H.; Zhang, K.; Liu, Q.; Lian, W. Low-rank weighted co-saliency detection via efficient manifold ranking. *Multimed. Tools Appl.* **2019**, *78*, 21309–21324. [CrossRef]
34. Tang, L.; Li, B.; Kuang, S.; Song, M.; Ding, S. Re-thinking the relations in co-saliency detection. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 5453–5466. [CrossRef]
35. Tan, Z.; Wan, L.; Feng, W.; Pun, C.M. Image co-saliency detection by propagating superpixel affinities. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 2114–2118.
36. Zhang, D.; Meng, D.; Han, J. Co-saliency detection via a self-paced multiple-instance learning framework. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 865–878. [CrossRef]
37. Han, J.; Cheng, G.; Li, Z.; Zhang, D. A unified metric learning-based framework for co-saliency detection. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 2473–2483. [CrossRef]
38. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
39. Luo, C.; Wu, Y.; Zhao, Y. SupCAM: Chromosome cluster types identification using supervised contrastive learning with category-variant augmentation and self-margin loss. *Front. Genet.* **2023**, *14*, 1109269. [CrossRef]
40. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
41. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
43. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
44. Yao, T.; Li, Y.; Pan, Y.; Wang, Y.; Zhang, X.P.; Mei, T. Dual vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, 1–13 . [CrossRef]