DCAN:DenseNet with Channel Attention Network for Super-resolution of Wireless Capsule Endoscopy

Abstract—**Wireless Capsule Endoscopy (WCE) captures images of the gastrointestinal (GI) tract and transmits the images in a wireless manner. Due to the hardware limitations of the capsule and the varying imaging conditions within the GI tract, the recorded images can have a low spatial resolution with a high frame rate or a high spatial resolution with a low frame rate. While it is general to have low spatial resolution to capture details of GI tract, low spatial resolution limits the detection of minor anatomical features and abnormalities in the small intestine and other portions of the GI tract. Super-Resolution (SR) is a class of software-based techniques that are used to enhance the resolution of a Low-Resolution (LR) image. This work proposes a new model referred as *DCAN*-DenseNet with Channel Attention Network for Super-resolution of LR WCE images. The design of *DCAN* consists of multiple strategies adopted from state-of-the-art methods such as Channel Attention Network (CAN) from RCAN and short dense connections from DenseNet to extract details from LR observation. Additionally, to improve the accuracy of the SR images, we create a derivative dataset of 10,000 images from a publicly available WCE dataset. The proposed approach has been validated against multiple state-of-the-art methods by conducting quantitative evaluation of perceptual metrics. The analysis is complemented with statistical validation to demonstrate the consistency of the proposed method over the other models for the SR task.**

Index Terms—**Wireless Capsule Endoscopy, Super Resolution, Channel Attention Network, DCAN**

I. Introduction

The Wireless Capsule Endoscopy (WCE) is a minimally invasive medical technology that utilizes a small, swallowable capsule equipped with a wireless camera to capture images and videos of GastroIntestinal (GI) tract. The captured video frames are transmitted to a recording device outside the patient's body. It allows for a comprehensive examination of the small intestine similar to conventional endoscopy but with additional convenience. The recorded images and videos provide valuable diagnostic information about GI disorders such as Crohn's disease, tumors, bleeding, or Inflammatory Bowel Disease (IBD) [1]. It generates an average of 50k to 60k images while moving through the GI tract, and a normal colon video test generates about 8 hours of RGB video data. Thus, the vast amount of data generated by WCE presents a challenge for medical professionals, who must verify numerous images or videos. Continued advancements in technology and image analysis algorithms further enhance the capabilities of WCE, leading to improved patient care and outcomes.

Resolution plays a crucial role in all vision-driven applications including medical diagnosis. A low resolution video/image can lead to wrong diagnostics for both machines and medical practitioners [2]. The image sensor equipped with High-Resolution (HR) can help in visualizing intricate details within the digestive tract, such as mucosal irregularities, ulcers, polyps, or early-stage tumors [3]. The clear and detailed images obtained from HR camera allowing doctors for targeted interventions or surgical procedures. However, a capsule consisting of an optical dome, illuminator, imaging sensor, battery, and RF transmitter in a capsule-shaped structure with a length of 26 mm and a diameter of 11 mm [1] can work in two modes. The small-sized structure leads to hardware limitations in terms of spatial resolution of sensor which is usually coarser. The minimum resolution obtained by a capsule used is 336×336 pixels with 24 frames per second (fps) [4] and the maximum resolution of 1 megapixel can reduce the frames rate to 5 fps. Having higher fps is advantageous in covering large area and despite of having numerous benefits of WCE technology, the operational fps suffers with inadequate frame resolution and video quality leading to adverse diagnostics [5]. Thus, there is a clear demand for methods capable of enhancing the resolution of capsule endoscopes to facilitate both subjective and objective analysis.

Image Super-Resolution (SR) is a software-driven method used to enhance LR image to its corresponding HR one. Single Image SR (SISR) and Multi-Image Super-Resolution (MISR) are the two types of SR methods, with SISR being more popular due to its advantages over the MISR, where multiple images of the same scene and image registration are required. However, SISR poses a challenging ill-posed problem as a single LR image may correlate to several HR solutions [6]. The recent advancement of deep learning techniques has resulted in number of techniques that can be used in SISR making it possible to use for other applications.

Inspired by the success of applications in other domains, we present a SR approach for WCE images using deep learning-based approach which we refer to as *DCAN*-DenseNet with Channel Attention Network. The proposed architecture incorporates the Channel Attention Network (CAN) mechanism for extracting high-level details by feature scaling in adaptive way. Such a mechanism allow us to leverage the high frequency details in WCE image to identify and retain abnormality present in the WCE images for downstream classification tasks like pathology classification. Additionally, we also introduce short skip connections to extract low-level features which are common in images from GI tract. The low level features are then combined with the high-level features to generate information-rich SR images. To enhance the reconstruction process and recover image details, we also employ bottleneck, deconvolution, and reconstruction layers. The potential of the proposed model is evaluated on a new derived dataset created from original Kvasir capsule endoscopy dataset [4]. Our contributions from this work are:

- A new SR approach that leverages Channel Attention Network (CAN) and Dense connections to generate SR images from LR images.
- The CAN in the proposed model adaptively re-scales the features by taking into account the inter-dependencies among different features. Further, the use of short skip connections in convolution layers specializes in excessing the high-level features to obtain low-level features in the input LR image, which is important for better detailing in reconstruction of SR image.
- Unlike other approaches, we propose training and testing of WCE images in Y-channel of $YCbCr$ which provides better performance metrics as compared to RGB color scheme and corresponds closely to human visual system (HVS). This is validated through empirically through various metrics such as Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) which asserts our intuition of using $YCbCr$ over RGB processing.
- Further, due to unavailability of datasets for SR tasks, we create a new derivative dataset from Kvasir Capsule Endoscopy [4] to train the SR network. The new dataset consists 10,000 samples which are manually pre-processed to improve the accuracy of the proposed network. All our experiments are conducted using state-of-the-art methods to demonstrate the applicability of proposed approach for SR generation and is supported by detailed analysis of various perceptual metrics.

II. Literature Review

The SR methods based on deep learning aim to capture the complex relationship between given LR and HR images. Dong et al. [7] introduced Super-resolution Convolutional Neural Network (SRCNN) consisting shallow network of having 3 layers. Later, Kim et al. increased the network depth to 20-layers proposed VDSR [8] and DRCN [9], which achieved notable improvements over the previous SRCNN indicating the importance of network depth. In similar lines, Lim et al. [10] further advanced this concept by creating the EDSR, a very wide network an exceptionally deep network consisting of approximately 165 layers, using simplified residual blocks. However, it is worth noting that merely stacking residual blocks to construct deeper networks does not necessarily lead to significant improvements.

Tong et al. introduced DenseNet [11] leveraging dense connections between convolution layers and growth rate to quantify the amount of new information added by each layer to the final reconstruction. Also due to dense connections all level of high, average and low level of features can be extracted easily. Thus, the DenseNet model utilizes feature maps from each layer that are merged with the previous layer, and the data is replicated multiple times for effective training of very deep networks. Zhang at el. proposed RCAN [12] model with Residual in Residual (RIR) structure where the Residual Group (RG) acts as the basic module and allows for residual learning in a coarse level through the use of Long Skip Connections (LSCs). This model also introduces a Channel Attention (CA) mechanism, which adaptively re-scales each channel-wise feature by modeling the inter-dependencies across feature channels. Such mechanism enables the network to focus on more useful channels, thereby enhancing its discriminative learning ability. A number of other SR works can have been proposed to enhance the perceptual quality of SR results. For instance, Ledig et al. [13] proposed an SRGAN model that improves the perceptual quality of super-resolved images beyond pixel-level improvements. Similarly, Wang et al. [14] proposed an Enhanced Super Resolution using GAN (ESR-GAN), which introduces several improvements over SRGAN. These works have been tested on visible (i.e., RGB scene) images and are also extended to medical data. Mahapatra et al. in [15] used Progressive GAN (P-GAN) for accurate detection and proper segmentation of anatomical landmarks on MRI images. Additionally, a few other SR techniques have also been utilized to improve the quality of images acquired by traditional endoscopic cameras. Yasin et al. [16] learned a mapping from low-to-high resolution mapping using conditional adversarial networks with a spatial attention block to improve the resolution by up to factors of ×8, ×10, ×12 respectively. However, the approach is limited to conventional endoscopy images. Thus, the super-resolution of WCE images is not attempted by researchers in the community to the best of or knowledge motivating us to focus on SR task for WCE images.

III. Proposed method: DenseNet with Channel Attention Network (DCAN)

With an aim to recover rich high-frequency details from capsule endoscopy images, the proposed approach consists design inspired from RCAN [12] and DenseNet [11]. The RCAN model is one of the state-of-the-art methods for SR of visible images which has introduced novel Channel Attention Network (CAN) to improve the learning ability of CNN network. Similarly, dense connections are usually employed in CNN network to learn effective features from LR images and also to reduce the effect of overfitting or underfitting. Motivated by these, we incorporated above concepts in the proposed method which we referred as *DCAN*-DenseNet with Channel Attention Network. In WCE images, low-frequency components display a relatively homogeneous pattern, the high-frequency elements typically correspond to regions characterized by edges, texture, and other intricate details. Thus, the use of CAN in the proposed model enhances the channel-wise feature representations, and hence, *DCAN* gains the advantage to extract information more precisely. The fusion of this with dense connections results in a powerful architecture that leverages the strengths of both RCAN and DenseNet, enabling it to effectively handle the task at hand and achieve superior performance in acquiring intricate details within WCE data.

The architecture of the proposed model for the task of SR of WCE images for upscaling factors $\times 4$ is depicted in Fig. 1. It can be observed that the WCE LR image is given to a convolution layer first to learn low-level features. After this, a single CAN layer is added to learn the features channel wise from the LR image. Subsequently, a series of DCAN blocks are employed to learn high-level features. Towards the end, the bottleneck layer is used to decrease the input feature maps and finally, the deconvolution layer is employed to upsample the feature images, and the output of reconstruction layer generates an SR image.

A. DCAN Block

In the proposed network, we utilize DCAN blocks as fundamental building unit. This design allows to enhance details and promotes feature reuse throughout the network, leading to more comprehensive and expressive representations at higher layers. The network architecture of DCAN block is depicted in Fig. 2. There are n number of DCAN blocks used in our architecture, which we fix to 8 empirically. Each DCAN block consists Channel Attention Network (CAN), one convolution layer and m number of DCAN layers (i.e., $m = 8$) that enable to extract high-level features in the output image. Moreover, one skip connection is added to avoid vanishing-gradient problem. The block schematic DCAN layer is displayed in Fig. 3 (a). Each DCAN layer consists of a convolution layer having kernel size 3×3 and Relu activation function with short skip connection. Thus, the proposed model consists short skip connections and also global skip connections for effective learning and also to avoid gradient problem.

B. Channel Attention Network (CAN)

The earlier CNN-based SR methods [7]–[11], [17], treat LR channel-wise features equally, which is not optimal for real-world cases. To address this issue and focus the network on more informative features, CAN mechanism is proposed in RCAN [12] that exploits the interdependencies among feature channels. Generating different attention for each channel-wise feature is a crucial step in this mechanism. An LR information contains both low-frequency and high-frequency components that are valuable for SR. However, the low-frequency parts are relatively homogeneous, while the high-frequency components typically correspond to regions with edges, texture, and other details. Second, each filter in the convolution layer operates within a local receptive field, which limits its ability to exploit contextual information beyond the local region. Thus, the use of CA in the proposed method is helpful to learn the features effectively by assigning proper weights to each feature. The architecture for CAN is depicted in Fig. 4. It consists of adaptive average pooling with a convolution layer having kernel size 3×3, attached with a ReLU activation function, which is passed to another convolution layer having kernel 1×1, and a skip connection is used to add the input values with the output of sigmoid function.

C. BottleNeck Layer

In order to enhance the compactness and computational efficiency of the model, we utilize a bottleneck layer to decrease the quantity of feature maps prior to their input into the deconvolution layers. The bottleneck layer shown in Fig. 3(b) is used to reduce the output features from DCAN blocks to a lower dimension. It consists of a convolution layer having kernel size 1×1 with ReLU activation function. In our proposed model, we are reducing the features to 256 features using the bottleneck layer.

D. Deconvolution and Reconstruction Layers

Deconvolution layers can be seen as the inverse operation of convolution layers, allowing for the learning of diverse upscaling kernels that work together to predict HR images. It provides two advantages: By conducting computations in the LR space, the SR reconstruction process is accelerated. Additionally, the inclusion of deconvolution layer enables the utilization of contextual information from LR images to infer high-frequency details. The network design of Deconvolution layer consists of two pixel-shuffle layers as shown in Fig. 3(c), where each layer upsamples the image by a factor of $\times 2$. Pixel-shuffle rearranges the feature maps by reshaping them into a higher resolution. It then rearranges the pixel values to get the final image. We are using two pixel-shuffle layers which give us total upsampling of factor 4. Finally, a reconstruction layer, consisting of a convolution layer with a 3×3 kernel, is used to generate SR images from the feature maps in the RGB space.

IV. Experimental Analysis

The design of the proposed model is validated by conducting subjective and quantitative evaluations. We empirically verify DCAN's effectiveness with state-of-the art architectures qualitatively by taking a patch from the output images from all state-of-the-art models. In addition, the same is verified quantitavively using different standard SR metrics such as Peak Signal to Noise Ratio (PSNR) & Structural Similarity Index Metric (SSIM) and using perceptual metric i.e., Learned Perceptual Image Patch Similarity (LPIPS). Finally, the statistical analysis of the proposed model are also presented in *Supplementary material* due to space constraints. Our method is benchmarked against state-of-the-art models such as SRGAN [13], CycleGAN [18], DenseNet [11], and RCAN [12] for comparison purposes.

A. Dataset

One of the novel contributions of the work is the creation of new derivative dataset from the available Kvasir Capsule Endoscopy Dataset [4] consisting of WCE images. In original Kvasir dataset, each image is in RGB color space with size of 336×336 pixels. The dataset contains a total of $47,236$ images, which are categorized according to different medical anomalies. As the original dataset contained redundant images with many border areas with black pixels, we have curated the dataset for the SR task by manually selecting images and removing redundant images from the Kavasir dataset. The new

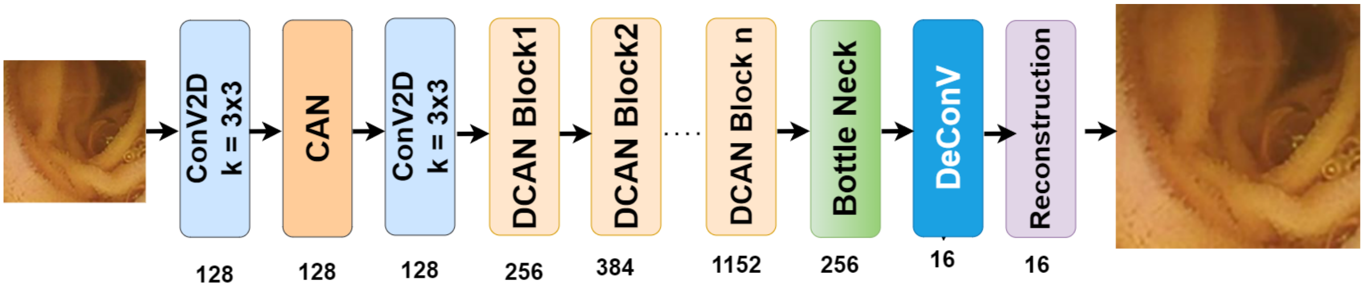

Fig. 1. The network architecture of proposed model *DCAN*, where k denotes kernel size and numerical values mentioned below every layer indicates size of output features.

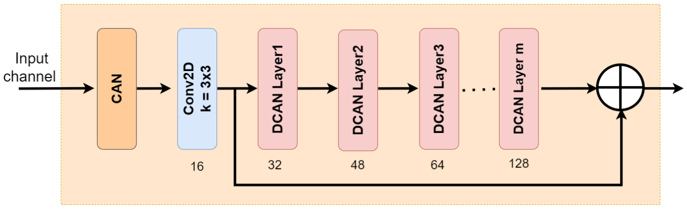

Fig. 2. The network architecture of DCAN block used in proposed model-*DCAN*. The values below every layer indicate the number of output features.

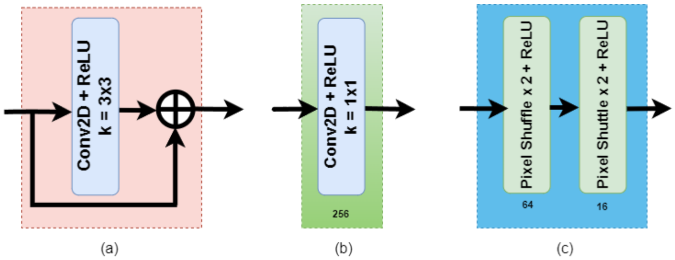

Fig. 3. The design of (a) DCAN Layer, (b) Bottleneck Layer and (c) Deconvolution Layer in the proposed method.

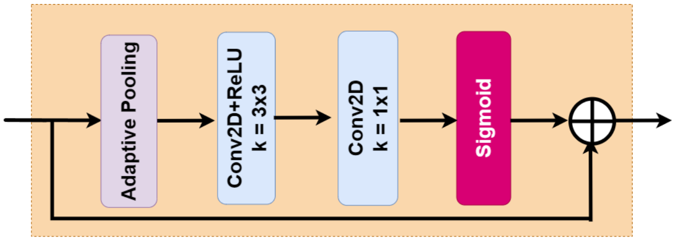

Fig. 4. The architecture design of Channel Attention Network (CAN) used in *DCAN* model.

SR dataset therefore consists of $10,000$ training images, 550 validation images and 1000 testing images [1]. As mentioned earlier, the WCE images from the original Kvasir dataset containing non-informative part in the border area. Those regions are removed manually through cropping resulting in images of 280×280 pixels for all images. The proposed model along with all other models are experimented on the new datset and SR results are generated.

B. Training details

Firstly, to prepare LR-HR pair of WCE images, we consider the original images as the HR image and applied bicubic down-sampling with factor ×4 and obtained LR image. These LR-HR pairs are fed to the proposed model to train it to generate SR images. Further, each LR image has also been transformed into $YCbCr$ space and only the Y-channel was used for training which represents a gamma-encoded channel that predominantly contains high-level feature information. On the other hand, the Cb and Cr channels are chroma-encoded channels that do not contain as many high-level features To save computational time during the process and improve the extraction of high-level features in the SR image, the Cb and Cr channels are directly interpolated and added to the output image. The training process aimed to minimize the loss function, which was taken as the Mean Squared Loss (MSE). The training was carried out for a total of 300 epochs with a batch size of 32. Additionally, the Adam optimizer was used with a learning rate of 0.0001. This protocol was used on all the state-of-the-art models and SR results are generated. While testing, we use YCbCr space of test LR image to generate SR image.

C. Comparison with state-of-the-art models

Qualitative Analysis: The qualitative comparison of various SR methods on scaling factor of ×4 is depicted in Fig. 5. One can inspect by looking at the zoomed-in patches that the proposed model generates better SR solutions than other models. Also, the SSIM map of each patch is shown below the patch SR image, which shows the similarity between generated SR and HR images. The yellow part in SSIM maps shows the similarity between SR and HR images. However, the green and blue regions in SSIM maps show the dissimilarity between SR and HR images. We can observe that the proposed model has more similar parts than the other methods. In the first row SSIM maps in Fig. 5, it can be observed that the bicubic output image exhibits the highest dissimilarity. Comparatively, other models such as RCAN and DenseNet perform better than other models as well as the bicubic method. However, the proposed model demonstrates the lowest dissimilarity among the bicubic method and all other state-of-the-art models. Additionally, in

[1] The dataset will be made available for researchers provide they have license agreement for original Kvasir Capsule dataset.

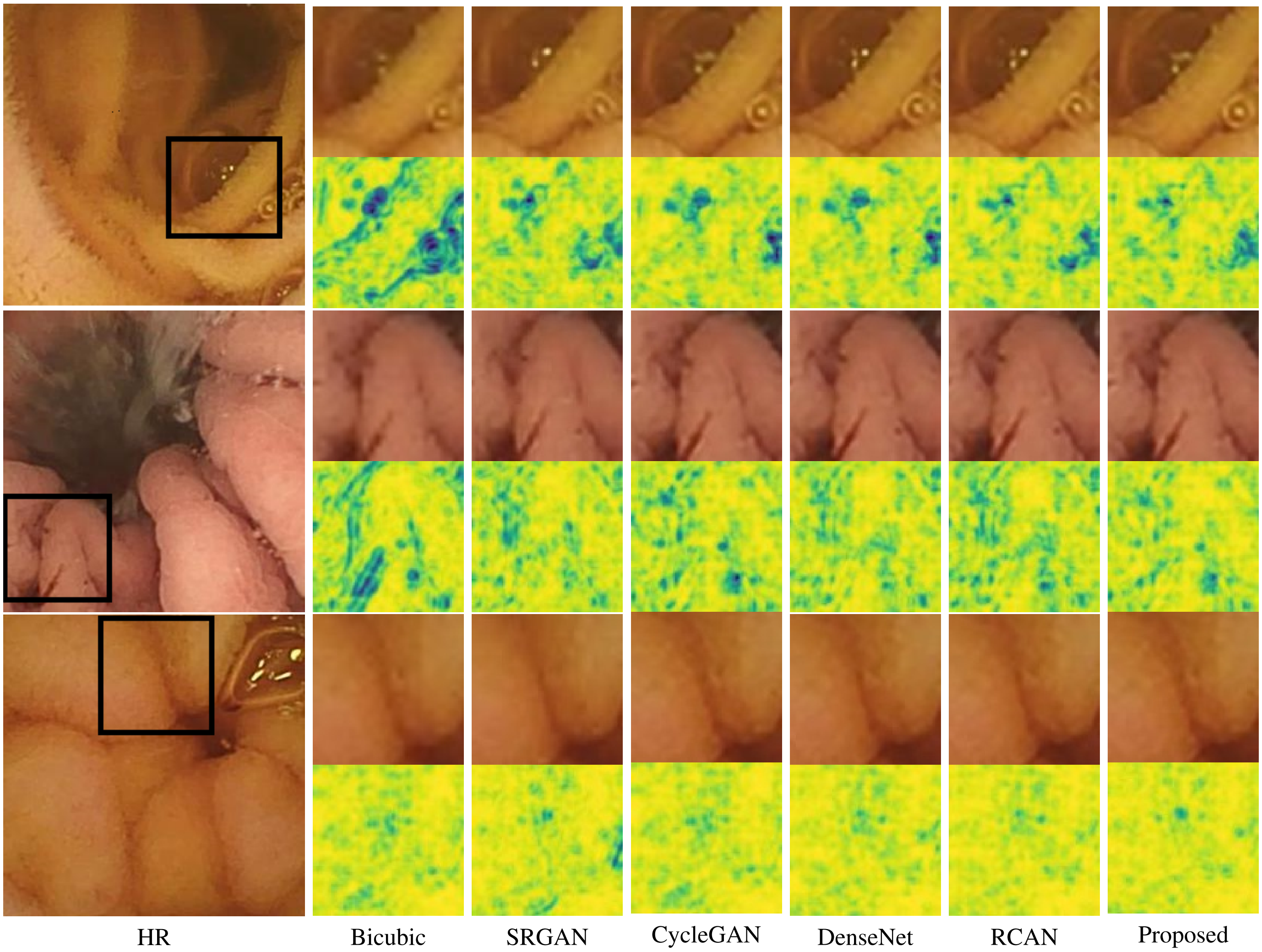

Fig. 5. Qualitative comparision of proposed models with state-of-the-art models using SSIM maps (where yellow region shows similarity and blue region shows disimilarity.)

second row, it is apparent that SRGAN and DenseNet exhibit better performance in comparison. However, the proposed model demonstrates the highest structural similarity, indicating superior performance in terms of preserving the structural characteristics of the image.

Quantitative Analysis: To validate the SR results quantitatively, the average SSIM and PSNR values for the testing images of each model are provided in Table I. We calculated the average PSNR and SSIM on Y channel as well as the RGB channels. From the table, it can be observed that the proposed DCAN model has the highest PSNR in both the Y channel and RGB channel, Additionally, the DCAN model also demonstrates the highest SSIM values, implying better structural similarity. When considering LPIPS for perceptual image comparison, lower LPIPS values emphasize the model's ability to capture perceptual similarity effectively. Remarkably, the DCAN model exhibits the lowest LPIPS values among all the different models, suggesting superior perceptual similarity.

TABLE I

QUANTITATIVE COMPARISON OF THE PROPOSED MODEL OVER OTHER MODELS USING DIFFERENT METRICS SUCH AS PSNR, SSIM AND LPIPS ON RGB AND Y-CHANNELS.

Model	PSNR ↑		SSIM ↑		LPIPS↓
	Y-channel	RGB	Y-channel	RGB	RGB
Bicubic	38.1069	37.2111	0.9296	0.9057	0.2310
SRGAN [13]	38.0377	37.0021	0.9291	0.9049	0.1972
CycleGAN [18]	38.0121	36.9441	0.9123	0.9012	0.1984
DenseNet [11]	39.6842	38.8596	0.9401	0.9369	0.1353
RCAN [12]	40.1438	39.4613	0.9427	0.9371	0.1359
Proposed	**40.2261**	**39.5389**	**0.9486**	**0.9378**	**0.1346**

Statistical Analysis: Finally, the statistical analysis is also conducted on the SR results of the proposed model along with the others to ensure the model consistency compared to state-of-the-art models. Standard deviation demonstrates the deviation in the values from the mean and hence it should be low for an algorithm. The values of standard deviation of each model are presented in Table II. As we can observe the values of our proposed model DCAN is lowest in comparision to all state-of-the-art models. From the given values, it can be noticed that while comparing PSNR consistency Y channel has

TABLE II
THE STATISTICAL COMPARISON OF THE PROPOSED MODEL WITH OTHER DIFFERENT METHODS USING PARAMETER OF STANDARD DEVIATION OF PSNR AND SSIM VALUES OVER MEAN VALUES IN RGB AND Y-CHANNELS.

Model	STD. dev. of PSNR ↓		STD. dev. of SSIM ↓	
	Y-channel	RGB	Y-channel	RGB
Bicubic	3.8943	2.2102	0.0353	0.0347
SRGAN [13]	3.3490	2.6924	0.0348	0.0324
CycleGAN [18]	3.6802	2.1937	0.0356	0.0319
DenseNet [11]	4.2025	3.1996	0.0378	0.0372
RCAN [12]	3.5612	2.8753	0.0367	0.0350
Proposed	**3.0006**	**2.0186**	**0.0270**	**0.0306**

lower consistency than RGB channels and while comparing SSIM, its vice-a-versa. Thus, one can conclude from this that there is high peaks in Y-channel i.e., Y-channel works great on majority images and provides better PSNR values, but as it focuses only on Y-channel, the features in Cb and Cr channels are not percieved properly, so when the high-level features are in Cb and Cr channels, it loses important information which is although a rare case as all important information is in Y-channel majorly. The box-plot representations of the same is discussed in *Supplementary material* due to space constraints.

V. CONCLUSION

Due to the hardware limitations of the WCE sensors, the captured data results in coarser resolution which affects the diagnosis accuracy of the diseases. We present a new SR approach *DCAN* using dense connections and channel attention modules to convert LR images to SR images. As the proposed network integrates the advantages of Channel Attention Network (CAN) from RCAN and utilizes short dense connections inspired by DenseNet, the proposed approach is able to effectively extract details from LR observations. Experiments show that the proposed network can perform better than other existing state-of-the-art SR models both quantitatively and qualitatively. The results are supported with a detailed analysis of quality assessment metrics such as PSNR, SSIM and LPIPS and statistical analysis of obatained results. A future direction in this work is to focus on improving the perceptual quality and assess it with medical practitioners.

REFERENCES

[1] P. Muruganantham and S. Balakrishnan, “A survey on deep learning models for wireless capsule endoscopy image analysis,” *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 83–92, 06 2021.

[2] S. B. and A. P., “Recent developments in wireless capsule endoscopy imaging: Compression and summarization techniques,” *Computers in Biology and Medicine*, vol. 149, p. 106087, 2022.

[3] O. Gilja, J. Hatlebakk, S. Ødegaard, A. Berstad, I. Viola, C. Giertsen, T. Hausken, and H. Gregersen, “Advanced imaging and visualization in gastrointestinal disorders,” *World journal of gastroenterology : WJG*, vol. 13, pp. 1408–21, 04 2007.

[4] P. H. Smedsrud, V. Thambawita, S. A. Hicks, H. Gjestang, O. O. Nedrejord, E. Næss, H. Borgli, D. Jha, T. J. D. Berstad, S. L. Eskeland, *et al.*, “Kvasir-capsule, a video capsule endoscopy dataset,” *Scientific Data*, vol. 8, no. 1, p. 142, 2021.

[5] C. F. Sabottke and B. M. Spieler, “The effect of image resolution on deep learning in radiography,” *Radiology. Artificial intelligence*, vol. 2, p. e190015, January 2020.

[6] H. Chen, X. He, C. Ren, L. Qing, and Q. Teng, “Cisrdcnn: Super-resolution of compressed images using deep convolutional neural networks,” *Neurocomputing*, vol. 285, 09 2017.

[7] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pp. 391–407, Springer, 2016.

[8] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, 2016.

[9] J. Kim, J. K. Lee, and K. M. Lee, “Deeply-recursive convolutional network for image super-resolution,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1637–1645, 2016.

[10] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144, 2017.

[11] T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in *Proceedings of the IEEE international conference on computer vision*, pp. 4799–4807, 2017.

[12] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in *Proceedings of the European conference on computer vision (ECCV)*, pp. 286–301, 2018.

[13] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114, 2017.

[14] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, “Esrgan: Enhanced super-resolution generative adversarial networks,” in *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.

[15] D. Mahapatra, B. Bozorgtabar, and R. Garnavi, “Image super-resolution using progressive generative adversarial networks for medical image analysis,” *Computerized Medical Imaging and Graphics*, vol. 71, pp. 30–39, 2019.

[16] Y. Almalioglu, K. Bengisu Ozyoruk, A. Gokce, K. Incetan, G. Irem Gokceler, M. Ali Simsek, K. Ararat, R. J. Chen, N. J. Durr, F. Mahmood, and M. Turan, “Endol2h: Deep super-resolution for capsule endoscopy,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4297–4309, 2020.

[17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[18] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, 2017.