Accelerating Super-Resolution and Visual Task Analysis in Medical Images

Ghada Zamzmi *,†, Sivaramakrishnan Rajaraman † and Sameer Antani

National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA;
sivaramakrishnan.rajaraman@nih.gov (S.R.); santani@mail.nih.gov (S.A.)
* Correspondence: alzamzmiga@nih.gov (G.Z.)
† These authors contributed equally to this work.

Abstract: Medical images are acquired at different resolutions based on clinical goals or available technology. In general, however, high-resolution images with fine structural details are preferred for visual task analysis. Recognizing this significance, several deep learning networks have been proposed to enhance medical images for reliable automated interpretation. These deep networks are often computationally complex and require a massive number of parameters, which restrict them to highly capable computing platforms with large memory banks. In this paper, we propose an efficient deep learning approach, called *Hydra*, which simultaneously reduces computational complexity and improves performance. The *Hydra* consists of a trunk and several computing heads. The trunk is a super-resolution model that learns the mapping from low-resolution to high-resolution images. It has a simple architecture that is trained using multiple scales at once to minimize a proposed learning-loss function. We also propose to append multiple task-specific heads to the trained *Hydra* trunk for simultaneous learning of multiple visual tasks in medical images. The *Hydra* is evaluated on publicly available chest X-ray image collections to perform image enhancement, lung segmentation, and abnormality classification. Our experimental results support our claims and demonstrate that the proposed approach can improve the performance of super-resolution and visual task analysis in medical images at a remarkably reduced computational cost.

---

1. Introduction

Super-resolution (SR) is a well-studied area of computer vision with several applications that range from surveillance imaging to medical imaging. It can be defined as the process of estimating a high-resolution (HR) image from one or multiple low-resolution (LR) images [1,2]. SR is categorized into a single image SR (SISR) and multiple images SR (MISR) based on the number of images used to reconstruct the HR image [1]. SISR recovers the HR image from a single LR image while MISR recovers the HR image from multiple LR images of the same object. MISR methods are less popular than SISR because it is difficult to find multiple LR images of the same object. In addition, MISR methods are less effective and have higher computational complexity because they require image registration and fusion prior to enhancement [1,2].

Medical imaging has been widely used in clinical practice to support accurate and real-time diagnosis [3]. Multiple imaging modalities, such as ultrasound (US), X-ray (XR), computed tomography (CT), and magnetic resonance imaging (MRI), are used to provide structural or functional information at different spatial and temporal resolutions. While several medical imaging modalities acquire images at a higher resolution, there are different situations where this is not feasible. Examples include the acquisition modality, lack of expertise, and reduced radiation exposure [3,4].

Several studies reported positive impacts of SR models on medical interpretation and diagnosis [2,5,6]. Specifically, it has been reported [5] that constructing HR images via SR models allows detection and recognition of subtle lesions, and reduce the misdiagnosis rates. In addition, it has been reported [2] that enhancing image resolution can often improve the performance of several subsequent tasks including detection, segmentation, and classification. The positive impact of SR models has been reported for specific imaging modalities. For example, several studies (e.g., [6]) demonstrated that applying SR models to chest X-ray (CXR) images can significantly improve the diagnosis of pulmonary disease. Using SR models to reconstruct HR CT from LR CT has also shown a significant improvement in diagnostic accuracy as reported in [7–9]. Further, Rundo et al. [10] reported the positive impact of using image enhancement models (MedGA [11]) on segmentation. In particular, it has been found that using the MedGA model [11] to enhance the resolution of MRI images increased the performance of uterine fibroid and brain tumor segmentation. Therefore, we believe the use of SR models to reconstruct HR images is of great significance in medical image analysis.

Existing methods for SISR are computationally complex, time intensive, and may require optimizing a large number of training parameters. Further, existing methods use individually trained models for SISR and visual task analysis. In this paper, we propose the *Hydra* approach for simultaneous SISR and visual task analysis. The proposed *Hydra* learns the mapping from LR to HR and uses several computing heads for visual task analyses. It reduces computational complexity while improving performance. The rest of the paper is organized as follows. Section 2 presents existing works for super-resolution and visual task analysis followed by a discussion of the contributions. We describe in Section 3 the CXR datasets used in this paper and present our proposed approach for SISR and visual task analysis. Section 4 presents the experimental setup followed by the results of our approach and a comparison with the state-of-the-art. Finally, we conclude the paper and discuss different directions for future research in Section 5.

2. Related Work and Contributions

In image processing, enhancement techniques focus on the visual improvement of the image quality while restoration techniques focus on restoring the image to its original quality. Super-resolution and contrast stretching are examples of image enhancement techniques while image in-painting is an example of image restoration. We present next existing methods for SISR and discuss the visual tasks that are typically applied to the output of SISR models.

2.1. Super-Resolution

Several SISR methods have been proposed to recover an HR image from a single LR image. These methods can be broadly divided into handcrafted methods and deep learning methods.

Early handcrafted SISR methods include interpolation-based methods, statistics-based methods, and example-based methods. Interpolation-based methods, such as bilinear [12], bi-cubic [13], and edge-directed interpolation [14], estimate HR by interpolating LR pixels to HR pixels. These types of methods tend to generate overly smooth or blurry image. Statistics-based methods (e.g., [15]) learns the statistical relationship of the gradient profile between the HR and LR images, motivated by the idea that shape statistics of the gradient profiles is invariant to the image resolution. These types of images tend to produce watercolor-like artifacts when applied to visually complex images. Finally, example-based methods use conventional machine-learning algorithms, such as Random Forests (RFs) [16–18] and Markov Random Fields (MRFs) [19], to learn the mapping from LR to HR images. These methods tend to generate a reconstructed image that contains irrelevant (hallucinated) details.

Recently, convolutional neural networks (CNNs) and generative adversarial networks (GANs) have been used to learn the mapping from LR to HR images. For example, Dong et al. [20] proposed the first SR-CNN network to efficiently learn this mapping. Kim et al. [21] proposed a CNN model, known as very deep super-resolution (VDSR), to predict the mapping from LR to HR images using residual

learning. Inspired by the VGG architecture [22], VDSR has 20 layers and is trained using extremely high learning rates. VDSR achieved state-of-the-art performance and outperformed SR-CNN [20]. It also resolved several issues of SR-CNN [20] such as the utilization of contextual information from a small image region, slow convergence, and the need for training individual scale-dependent models.

A GAN-based network, named P-SRGAN, has been proposed [23] recently to learn the mapping from LR to HR images. P-SRGAN consists of a generator network and a discriminator network. The generator network takes an LR image as input and generates the HR image while the discriminator compares the generated image with the HR image to generate good quality reconstructed images. Although this approach achieved excellent performance, GAN-based networks are hard to train due to the Nash equilibria problem. This problem is defined as a zero-sum game between the generator network (player 1) and the discriminator network (player 2) where the opponent players contest with each other in a game to improve their objective functions [24]. In addition, GAN-based networks are highly sensitive to the hyperparameter selection and often get into mode collapse (i.e., the generator maps different inputs to the same output) [24]. To regularize the training of GAN-based SR models and enforce the right mapping between the input and output domains, You et al. [25] proposed GAN-CIRCLE network for constructing HR CT images from their LR counterparts. The proposed network combines four loss functions, viz., adversarial loss, cycle-consistency loss, identity loss, and joint sparsifying transform loss, to stabilize training and enforce the right input-output mapping. Expert radiologists evaluation of GAN-CIRCLE on three CT datasets demonstrates its ability in constructing HR images from noisy LR input images. Although the usage of GAN-CIRCLE solved regular GANs training and mapping problems, this network is computationally complex and requires a relatively large GPU as well as much longer training time in comparison to the regular-GAN networks. Further, the network failed to faithfully recover subtle structures in CT images as discussed in [25]. We refer the reader to [26,27] for comprehensive reviews of other handcrafted and deep learning SISR methods.

While existing deep learning models achieved excellent performance and successfully constructed HR images, these models are trained using a single scale (e.g., [20,25]), have a deep architecture with a very large number of training parameters (e.g., [21,23,25]), and are hard to train (e.g., [23]). To resolve issues with existing models, we propose a simple SISR model that learns the mapping from LR to HR images. Our customized SISR model has the following virtues:

- **Simplicity and Stability.** Our customized SISR model, inspired by VDSR (20 conv. layers) [21], has a shallower structure (7 conv. layers) with a lower number of training parameters. In addition, our proposed SISR model is easy to train and has a single network contrary to GAN-based networks which are difficult to train and require both generator and discriminator networks. Further, our proposed model is more stable and less sensitive to hyper-parameters selection as compared to most GAN-based models. As large models with massive number of parameters are restricted to computing platforms with large memory banks and computing capability, developing smaller and stable networks without losing representative accuracy is important to reduce the number of parameters and the storage size of the networks. This would boost the usage of these networks in limited-resource settings and embedded healthcare systems.

- **Multiple Scales Training:** Our SISR model is trained with different scale factors at once. The trained network can then be tested with any scale used during training. As discussed in [21], training a single model with multiple scale factors is more efficient, accurate, and practical as compared to training and storing several scale-dependent models.

- **Context:** We utilize information from the entire image region. Existing methods either rely on the context of small image regions (e.g., [20]) or large image regions (e.g., [21]), but not the entire image region. Our experimental results demonstrate that using the entire image region leads to better overall performance while decreasing computations.

- **Raw Image Channels:** We propose to compute the residual image from the raw image (RGB or grayscale) directly instead of converting the images to a different color space (e.g., YCbCr [21]).

The residual image is computed by subtracting the HR reference image from the LR image that has been upscaled using interpolation to match the size of the reference image. The computed residual image contains information of the image's high-frequency details. The main benefit of directly working on the raw color space is that we decrease the total computational time by dropping two operations: (1) converting from raw color space to another color space (e.g., YCbCr) and (2) converting the image back to its original color space. Our customized SISR model computes the residual images directly from the original color space and learns to estimate these images. To construct an HR image, the estimated residual image is added to the upsampled LR image.

- **Combined Learning Loss:** We propose to train the proposed SISR model using a loss function that combines the advantages of the mean absolute error (MAE) and the Multi-scale Structural Similarity (MS-SSIM). Our experimental results show that MAE can better assess the average model performance as compared to other loss metrics. Also, our experimental results show that the MS-SSIM preserves the contrast in high-frequency regions better than other loss functions (e.g., SSIM). To capture the best characteristics of both loss functions, we propose to combine both loss terms (MAE + MS-SSIM).

In summary, our proposed SISR model has a simple architecture, is stable, and it is trained using the entire raw image and multiple scales at once to minimize a proposed loss function. It reduces the number of training parameters and storage size without performance degradation. Such advantages would boost the usage of SISR models in limited computational resources and facilitate the deployment of these models in clinical settings for potential real-time healthcare applications.

2.2. Visual Task Analysis

Existing methods for medical image analysis apply SISR models to construct HR images followed by using the constructed HR images as input to individual models for individual tasks. For example, various methods applied deep learning-based SISR models to LR images and used the constructed HR images as input to separate segmentation (e.g., U-Net network [23,28,29]) and classification (e.g., VGG network [23,28] and DenseNet network [30]) models. This traditional approach for analysis is not efficient because it involves unnecessary repetitions of learning (end-to-end) multiple task-specific models in isolation.

Contrary to previous works that separate image enhancement from other visual tasks, we propose to use our customized SISR model as a shared representation to simultaneously learn multiple subsequent visual tasks. Specifically, the weights of our SISR model, which learns the mapping from LR to HR, are directly used to simultaneously learn tasks such as image segmentation and classification. Using the proposed SISR model as a shared backbone improves generalization and prevents unnecessary repetitions of learning visual task models in isolation. This can lead to a decrease in resource utilization and training time, and boosts the use of deep learning models in limited-resource settings.

2.3. Contributions

In this paper, we propose the *Hydra*, a deep learning approach that consists of two components: a shared trunk and computing heads. The trunk is a customized SISR model that learns the mapping from LR to HR. The trained trunk is then appended with task-specific layers to learn multiple visual tasks in medical images. Figure 1 depicts the main difference between the *Hydra* and existing works. As can be seen from the figure, the *Hydra* trunk is used as a shared backbone to learn multiple visual tasks (heads). On the contrary, the majority of existing methods (traditional approach for task analysis) use the constructed HR image as input to multiple individual models. The main contributions of this paper can be summarized as follows:

- We propose the *Hydra* approach for enhancing medical image resolution and visual task analysis. The *Hydra* consists of two components: a shared trunk and computing Heads.

- *Hydra* trunk is a proposed customized SISR model that learns the mapping from LR to HR. This SISR model has a simple architecture and is trained using the entire raw image and multiple scales at once to minimize a proposed loss function. Our experimental results show that the proposed SISR model, which has a markedly lower number of training time and parameters, achieves state-of-the-art performance.
- We propose to append the customized SISR trunk with multiple computing heads to learn different visual tasks in medical images. We evaluate our approach using CXR datasets to generate HR representation followed by jointly performing lung segmentation (visual task 1) and abnormality classification (visual task 2). We focus mainly on these two tasks because classification and segmentation are the key tasks in most medical image analysis applications.
- We empirically demonstrate the superiority and efficiency of our approach, in terms of performance and computation, for SR and medical image analysis as compared to the traditional approach.

We present next the CXR datasets used to evaluate the proposed *Hydra* and provide detailed descriptions of *Hydra* trunk and computing heads.
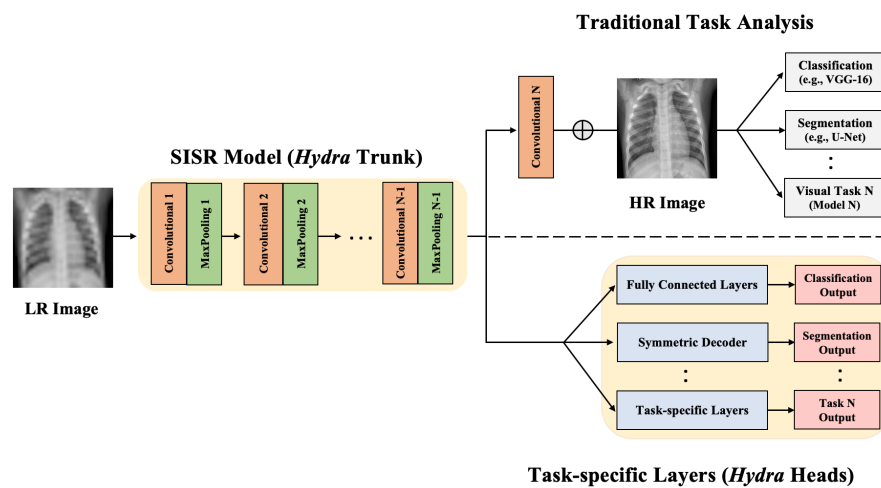

Figure 1. Illustration of the traditional approach and proposed *Hydra* for image enhancement and visual task analysis. The SISR model (*Hydra* trunk) is used to learn the mapping from LR to HR. In the traditional approach, the constructed HR image is used as input to different models; e.g., segmentation and classification models. In the proposed *Hydra*, the learned SISR model is used as a shared trunk and appended with task-specific heads. The detailed architecture of *Hydra* trunk is shown in Figure 2.
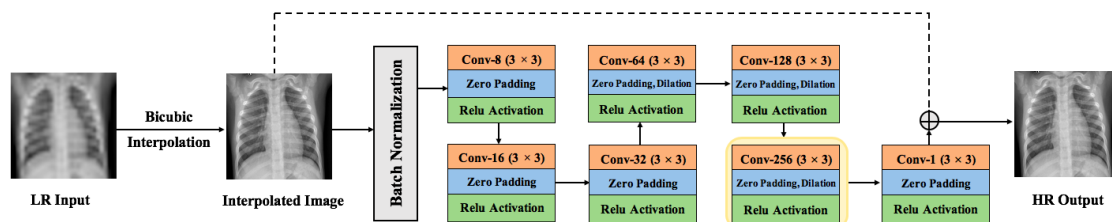

Figure 2. The structure of *Hydra* trunk (SISR model). To learn subsequent visual tasks, we truncated the SISR model at the deepest convolutional layer (yellow box) and appended task-specific heads as shown in Figure 1.

3. Materials and Methods

3.1. Datasets for Training and Testing

We used four publicly available CXR datasets: Radiological Society of North America (RSNA) [31], Shenzhen [32], Montgomery [32], and Japanese Society of Radiological Technology (JSRT) [33] CXR collections. The RSNA CXR dataset is used to build the customized SISR model (trunk) with 70%, 20%, 10% patient-level splits for training, validation, and testing, respectively. The Shenzhen, Montgomery, and JSRT datasets are used alternately for training and testing the task-specific heads.

3.1.1. RSNA CXR

The RSNA dataset contains 30,000 CXR exams, in which 15,000 had positive findings for pneumonia-related pulmonary opacities. The remaining 15,000 negative exams are taken from two groups: 7500 exams had no findings while the remaining had pathologies unrelated to pneumonia. The dataset is provided as DICOM files and has a total of 26,684 frontal images. These images have 1024×1024 spatial resolution. The ground truth annotations are provided, for CXRs containing pneumonia-related opacities, as bounding boxes using a commercial annotation system. These ground truth boxes are annotated by six board-certified radiologists. Further details of the RSNA CXR dataset can be found in [31].

3.1.2. Shenzhen CXR

The Shenzhen dataset is collected as a collaboration between the Shenzhen People's Hospital and the Guangdong Medical College, Shenzhen, China. CXRs are collected from outpatient clinics within one-month period, mostly in September 2012, using a Philips DR Digital Diagnostic system. This dataset contains 662 frontal CXR, of which 326 are normal cases and 336 are cases with manifestations of tuberculosis (TB). All the images are provided in Portable Network Graphics (PNG) format. The size of the images in this collection varies, but it is approximately 3000×3000 pixels. Each image is provided with a clinical reading text file that contains the patient's age and gender as well as normal and abnormal labels. In addition, each image has a corresponding binary ground truth mask, generated by manually segmenting lung regions. Further details of the Shenzhen CXR dataset can be found in [32].

3.1.3. Montgomery CXR

The Montgomery dataset is collected as part of the collaboration between the Montgomery County (Maryland, USA) and the Department of Health and Human Services. The dataset contains 138 frontal chest X-rays from the Montgomery County's Tuberculosis screening program. The CXR collection of this dataset is acquired using the Eureka stationary CXR machine. Eighty cases are normal and the rest are cases with TB manifestations. All the images are provided in PNG format with either 4020×4892 or 4892×4020 spatial resolution. Each image is provided with a clinical reading text file that contains the patient's age and gender as well as normal and abnormal labels. In addition, each image has a corresponding binary ground truth mask, generated by manually segmenting lung regions. The manual lung segmentation was performed under the supervision of a radiologist, following anatomical landmarks, such as the boundary of the heart, pericardium, and aortic arc. Further details of the Montgomery CXR dataset can be found in [32].

3.1.4. JSRT CXR

The JSRT dataset was collected by the Japanese Society of Radiological Technology (JSRT) in cooperation with the Japanese Radiological Society (JRS). This dataset contains 247 CXR images, 154 with lung nodules and the rest without lung nodules. All the CXR images have 2084×2084 size with spatial resolution of 0.175 mm/pixel and 12-bit grayscale color depth. Each image is provided with

a text file that contains additional information such as the patient age, gender, diagnosis (malignant or benign), and coordinates of the nodule. We used the binary masks, which are available at [34], as the ground truth masks for segmentation. Further details of the JSRT CXR dataset can be found in [33].

3.2. Proposed Approach: Hydra

We present next both components of the *Hydra*: the SISR trunk and task-specific heads.

3.2.1. Hydra SISR Trunk

The structure of our SISR model is outlined in Figure 2. As shown in the figure, the model has a shallower depth [21]. The first layer of the network operates on the entire input image (RGB/grayscale), contrary to previous methods [20,21]. This input image is interpolated, via bi-cubic interpolation, and fed into the batch normalization layer. This layer optimizes and speeds up model training, and improves generalization. Our proposed SISR model has seven convolutional layers with 3×3 filter size. The number of filters starts as 8 in the first layer and continues to double in subsequent layers. We used zero padding to (1) keep the sizes of all feature maps the same and (2) avoid boundary artifacts due to the convolution operation. We also used ReLU functions with all convolution layers to speed-up model training, resulting in faster convergence. The last three convolutional layers (4th , 5th, and 6th) of the SISR model have dilated convolutional kernels (dilation rate = 2) that capture a wider context, by enclosing a bigger receptive field, at a reduced computational cost. To construct an HR image, the SISR model takes an interpolated LR image as input and predicts the residual image, which contains high-frequency details. As discussed in [21], predicting residual image, instead of non-residual image, leads to faster convergence and superior performance. Finally, the predicted residual image is added back to the input LR image to get the final HR image.

We trained the model using pairs of (LR,HR) CXR data. Specifically, we used bi-cubic interpolation to down-sample the HR images for training the model following existing methods in the literature [20,21,23]. We trained the SISR model using Adam optimizer with 1×10^{-3} learning rate, 256 epochs, 64 batch size, and random weight initialization. The model is trained on multiple scales (i.e., 2, 3, 4, and 8) at the same time. Training is carried out to find the optimal parameters by optimizing the learning-loss function, which consists of two terms: MS-SSIM and MAE. The combined learning-loss achieved the best results as it allows to capture the best characteristics of both loss functions. Depending on the number of GPUs, the training time of our SISR model ranges from 21,917.64 to 54,311.21 seconds.

3.2.2. Hydra Task-Specific Heads

The traditional approach in medical image analysis involves using separate deep learning models to perform different visual tasks. As shown in Figure 1, a deep learning model would be used to construct HR images from LR images. The constructed HR images are then used as inputs to different individual models for different visual tasks. This traditional approach for analysis is not efficient because it involves unnecessary repetitions, has high resource utilization, and requires a relatively large amount of annotated data.

In this paper, the trained SISR model (trunk) is used as a shared backbone to jointly solve any number of target tasks that use CXR imaging modality; i.e., the target tasks operate on the same imaging modality or have similar input distributions. To achieve this, we truncated the SISR model at the deepest convolutional layer (6th convolutional layer) and appended task-specific heads. Specifically, the optimized weights of SISR are used to transfer modality-specific knowledge and serve as a promising initialization that can be transferred and repurposed for the target tasks. To learn task-specific features, task-specific layers are appended to the shared SISR model. We believe this approach is suitable for medical images because visual medical tasks often operate on the same input modality and have different outputs.

4. Experiments and Results

In this section, we evaluated the performance of *Hydra* using four publicly available CXR datasets (see Section 3.1). We compared the performance of the *Hydra* trunk (SISR model) with VDSR [21]. We also compared the performance of *Hydra* heads with the traditional approach (individual models for different visual tasks) and the state-of-the-art methods. The code of this work and all the training parameters can be found in: Hydra Github Page.

4.1. Hydra Trunk Performance

We evaluated the performance of our customized SISR model using the combined loss (MAE + MS-SSIM), peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and MS-SSIM evaluation metrics. Various studies [35,36] found that SSIM provides a better indication of SISR performance than PSNR because it can better represent human visual perception. Also, it has been reported [35,36] that MS-SSIM is better than SSIM because it is calculated over multiple scales through a process of multiple stages of sub-sampling. We used the RSNA CXR dataset with 70%, 20%, and 10% patient-level splits for training, validation, and testing, respectively. We downscaled all the images ($256 \times 256 \times 1$) by scale factor 2, 3, 4, and 8 using bi-cubic interpolation. We used interpolation following existing methods in the literature [20,21,23]. We implemented our SISR model using Keras [37] Python library with TensorFlow backend.

Table 1 shows the results of our customized SISR model and VDSR on the testing set. Because both models are trained on multiple scales (2, 3, 4, and 8) at once, they can be tested with any scale used during training. As can be seen in the table, our SISR model outperformed VDSR in most cases. Recall that VDSR, which contains 20 convolutional layers, is a large model with a relatively large number of parameters as compared to our model (7 layers). These results suggest that our customized SISR can faithfully and efficiently construct HR images. Figure 3 shows examples of SR results with scale factors of 2, 3, 4, and 8 using the *Hydra* trunk.

Table 1. Performance of the proposed SISR model (*Hydra* trunk) and VDSR.

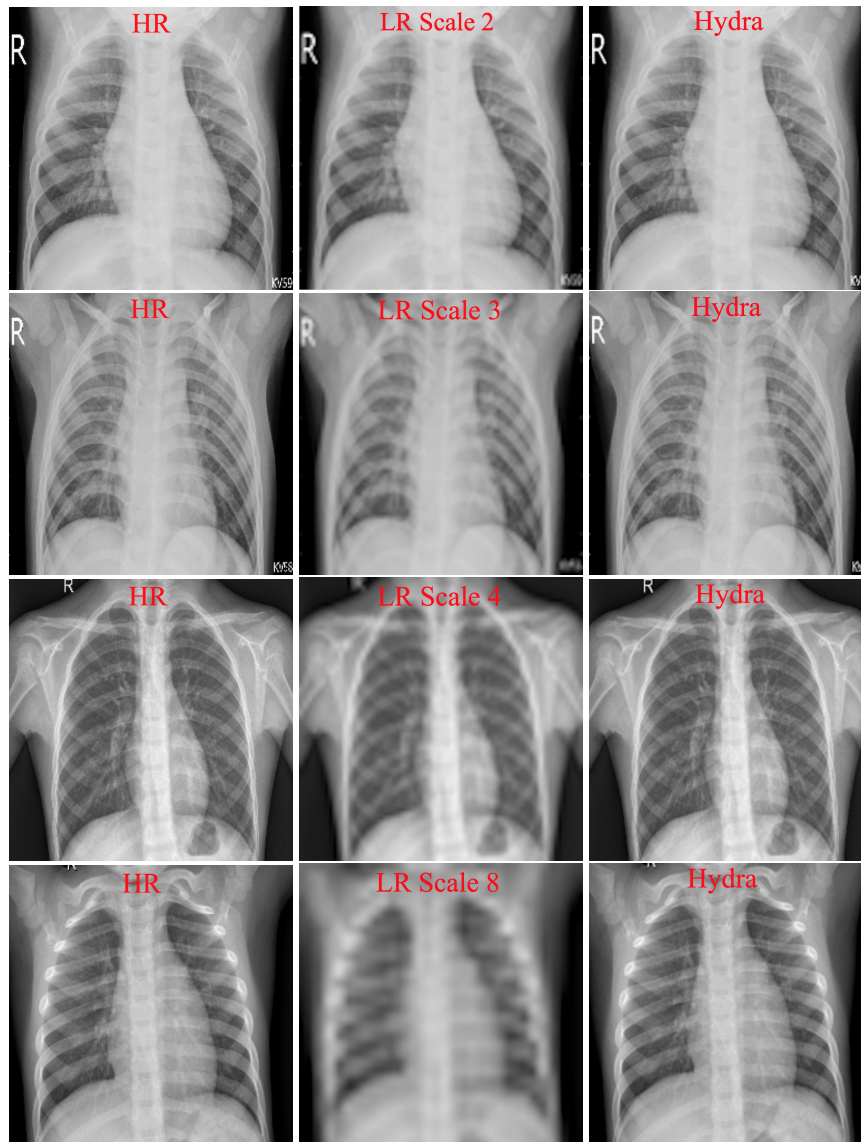| Scale	Model	Proposed Loss	PSNR	SSIM	MS-SSIM
2	Proposed	0.0072	34.822	0.9366	0.9911
	VDSR	0.0091	32.564	0.9355	0.9801
3	Proposed	0.0161	32.274	0.9000	0.9861
	VDSR	0.0157	31.972	0.8858	0.9815
4	Proposed	0.0201	30.764	0.8706	0.9745
	VDSR	0.0250	30.691	0.8652	0.9722
8	Proposed	0.0454	27.306	0.7839	0.9233
	VDSR	0.0478	25.087	0.7644	0.9088

Figure 3. Examples of reconstructed SR images. First column: original HR images. Second column: downsampled input LR images. Third column: reconstructed images obtained using the proposed SISR model (*Hydra* Trunk).

4.2. Hydra Heads Performance

The trained SISR model (*Hydra* trunk) can be used as a shared backbone to solve any number of target tasks that operate on CXR modality. To solve a new visual task, we only need to append the shared SISR model with task-specific layers followed by training these layers. In this paper, we experimented with segmentation and classification because they represent the key tasks in most medical image analysis applications [38]. However, the *Hydra* is flexible and can be easily extended to include any number of tasks (e.g., detection). We describe next task-specific heads for CXR abnormality classification and lung segmentation. In all experiments, the Shenzhen, Montgomery, and JSRT datasets are used alternately for training and testing the task-specific heads.

4.2.1. Abnormality Classification

We define the abnormality classification task as categorizing CXR images into normal or abnormal. The abnormal class contains CXR images with different types of pulmonary opacities caused due to TB or lung nodules. Deep learning approaches have been widely used for lung abnormality classification.

For example, Li et al. [30] presented a patch-based multi-resolution CNNs for CXR lung nodule classification. Prior to deep feature extraction, the CXR images are enhanced using a histogram-based method. After enhancing the images, three multi-resolution patches are extracted from each lung field and used to train three CNNs at different resolutions. The final classification label is then generated by fusing the outputs of the three CNNs. The proposed method achieved 0.98 Free-Response Receiver Operating Characteristic (FROC) curve when evaluated on the JSRT dataset. In case of CT images, Han et al. [39] proposed a GAN-based network, called 3D MCGAN, to generate synthetic and diverse nodules in lung CT scan images at a desired location. The generated synthetic images are then used by 3D Faster RCNN as additional training data for lung nodule detection. The experimental results suggest that the proposed method can achieve higher sensitivity and overcome the medical data paucity. Contrary to these works, we propose to simultaneously learn pre-processing operation and CXR abnormality classification using a single trunk and a task-specific head.

To create a task-specific head for abnormality classification, we initialized the SISR model with its weights, freezed the weights, and appended task-specific layers to the truncated SISR model (at the deepest or 6th convolutional layer). The task-specific layers are global average pooling (GAP), fully connected (FC), dropout (D), and Softmax (SM) layers. We then trained task-specific layers on the Shenzhen dataset to minimize the categorical cross-entropy (CCE) loss. We trained the classification layers using Adam optimizer, 1×10^{-3} learning rate, 32 epochs, and a batch size of 8.

As a baseline, we used a model that has six convolutional layers followed by GAP, FC, D, and SM layers. The baseline classification model has the following architecture: input (256×256) → Conv2D (16) → MaxPool → Conv2D (32) → MaxPool → Conv2D (64) → MaxPool → Conv2D (128) → MaxPool → Conv2D (256) → MaxPool → Conv2D (512) → GAP → FC → D → SM. We randomly initialized the model's weights and optimized for the search parameters. We trained the baseline classification model to minimize the CCE loss. We performed two baseline experiments using the baseline classification model. In the first experiment, we used the original HR images as input to the classification model. In the second experiment, we applied bi-cubic interpolation to generate interpolated images and used the generated interpolated images as input to the baseline classification model. Recall that the main difference between the proposed approach and the traditional approach is that our approach directly uses the SISR model and only learns the task-specific classification layers.

Table 2 presents the performance of abnormality classification using the baseline approach and the proposed approach (*Hydra* classification head). We trained the head and the baseline model on the Shenzhen dataset, and evaluate their performance on the Montgomery dataset. The performance was reported using the accuracy, AUC, sensitivity, precision, and F-score. As shown in the table, the *Hydra* outperformed baseline models in all scales. These results suggest that freezing the weights of the trunk while learning the weights of the task-specific head can lead to better overall performance as compared to the traditional approach. Specifically, using the weights of the SISR model improves generalization and decreases the number of training parameters (see Section 4.3) for the classification head, which leads to enhanced overall performance.

Table 2. Performance of CXR abnormality classification using the baseline model and *Hydra* head.

Model	Accuracy	AUC	Sensitivity	Specificity	Precision	F-Score
Baseline (HR)	0.8283	0.8502	0.7800	0.8776	0.8667	0.8211
Baseline (scale 2)	0.8182	0.8586	0.7600	0.8561	0.8636	0.8085
Hydra Head (scale 2)	0.8512	0.8905	0.7700	0.9082	0.8953	0.8280
Baseline (scale 3)	0.7980	0.8492	0.6800	0.8475	0.8947	0.7727
Hydra Head (scale 3)	0.8384	0.8748	0.7500	0.9184	0.9000	0.8000
Baseline (scale 4)	0.7929	0.8632	0.6900	0.8980	0.8734	0.7709
Hydra Head (scale 4)	0.8219	0.8416	0.7100	0.9388	0.9189	0.7816
Baseline (scale 8)	0.7727	0.8537	0.7000	0.8469	0.8041	0.7568
Hydra Head (scale 8)	0.8081	0.8608	0.7800	0.8761	0.8235	0.7991

4.2.2. Lung Segmentation

To create a task-specific head for lung segmentation, the trained SISR model was truncated at the deepest convolutional layer and a symmetrical decoder was constructed and appended. Specifically, the SISR model (*Hydra* trunk) was truncated at the sixth layer in Figure 2 and appended with the following layers: Conv2D (128) → Conv2D (64) → Conv2D (32) → Conv2D (16) → Conv2D (8) → Conv2D (1). The segmentation head was trained to minimize the binary cross-entropy (BCE) loss. We used the weights of the SISR model to train the segmentation layers. We trained the segmentation layers using Adam optimizer, 1×10^{-3} learning rate, and 200 epochs.

As a baseline, we randomly initialized an individual model, which has the same encoder-decoder structure as the segmentation head, and optimized it for the search parameters. The baseline segmentation model has the following architecture: input (256×256) → Conv2D (8) → Conv2D (16) → Conv2D (32) → Conv2D (64) → Conv2D (128) → Conv2D (256) → Conv2D (128) → Conv2D (64) → Conv2D (32) → Conv2D (16) → Conv2D (8) → Conv2D (1). We trained the baseline segmentation model to minimize the BCE loss. We performed two baseline experiments using the baseline segmentation model. In the first experiment, we used HR images as input to the baseline segmentation model. In the second experiment, we applied bi-cubic interpolation and used the interpolated images as input to the baseline segmentation model.

Table 3 presents the performance of lung segmentation using the proposed approach and baseline model. Both models were trained on the Shenzhen dataset and evaluated on the Montgomery dataset. The performance was reported using the pixel accuracy, loss, and intersection over union (IoU). As shown in the table, the segmentation head, which uses the trunk as a backbone, outperformed baseline models in most scales. Figure 4 presents testing examples of the segmentation generated by the segmentation head as well as baseline approach and ground truth masks. The figure shows a clear visual improvement in the segmentation obtained by our proposed approach as compared to the traditional approach. These results suggest that initializing the segmentation head with the weights of the trunk enhances the overall performance. Recall that *Hydra* head uses the trunk weights while the baseline model is initialized randomly.

Table 3. Performance of *Hydra* segmentation head and baseline model on Montgomery dataset.

Model	Accuracy	Loss	IoU
Baseline (HR)	0.9863	0.0355	0.9464
Baseline (scale 2)	0.9700	0.0545	0.9263
Hydra Head (scale 2)	0.9858	0.0316	0.9516
Baseline (scale 3)	0.9613	0.0496	0.9285
Hydra Head (scale 3)	0.9897	0.0248	0.9599
Baseline (scale 4)	0.9390	0.0557	0.9201
Hydra's Head (scale 4)	0.9801	0.0325	0.9229
Baseline (scale 8)	0.9516	0.0816	0.8839
Hydra Head (scale 8)	0.9830	0.0411	0.9391

In addition to the Shenzhen and Montgomery datasets, we alternated between the Shenzhen, Montgomery, and JSRT datasets for training and testing to further evaluate the performance of the segmentation head. The testing sets of the Shenzhen and JRST datasets contain 100 and 140 images, respectively. Table 4 presents extensive evaluations of *Hydra* segmentation head. These results suggest the excellent performance of the proposed approach in well-established datasets. Table 5 compares the performance and computation time for lung segmentation using the proposed *Hydra*, anatomical atlas method [40], and efficient U-Net [41]. As can be seen in the table, the proposed *Hydra* outperforms similar approaches in the literature with the minimum computation time per image.
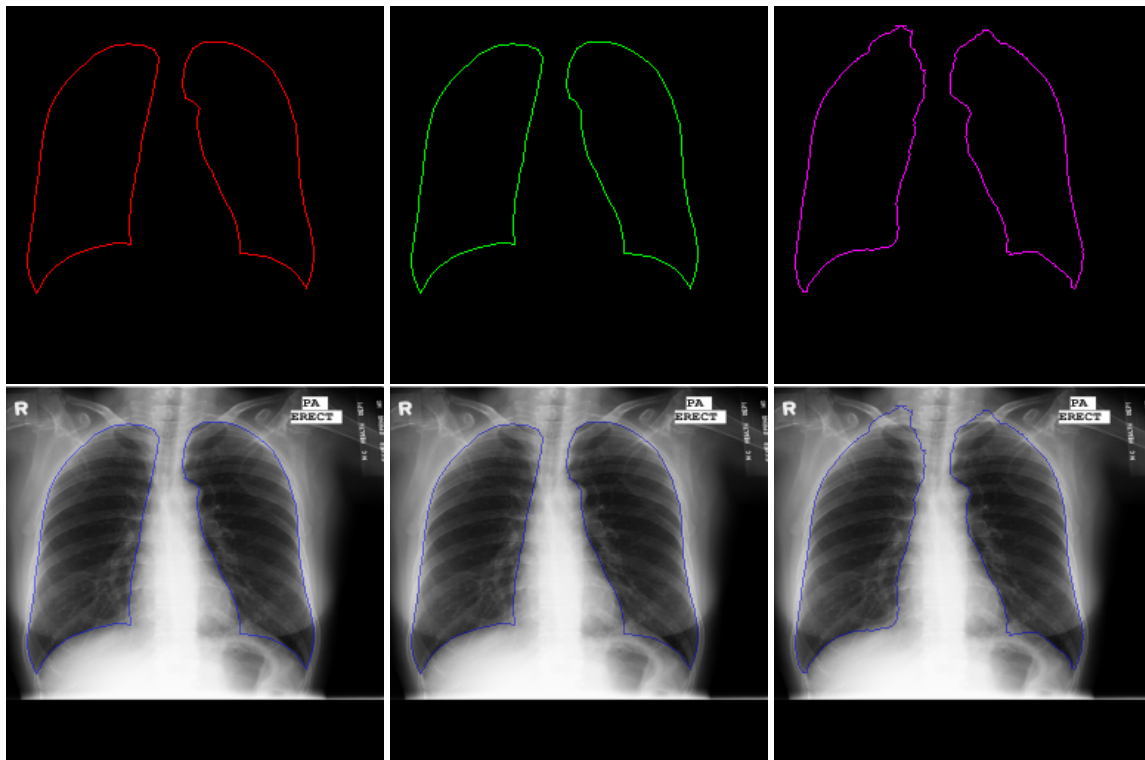
Figure 4. Example of lung segmentation. 1st column: ground truth red contour and the superimposition on the original image. 2nd column: green contour obtained by *Hydra* segmentation head and the superimposition on the original image. 3rd column: purple contour obtained by baseline (traditional approach) and the superimposition on the original image.

Table 4. Performance of *Hydra* head for CXR lung segmentation on different training and testing sets.

Hydra Scale	Train	Test	Accuracy	Loss	IoU
2	Shenzhen	Shenzhen	0.9762	0.0125	0.9521
	Shenzhen	JSRT	0.9832	0.0351	0.9265
	Montgomery & JSRT	Shenzhen	0.9768	0.0561	0.9294
3	Shenzhen	Shenzhen	0.9647	0.0297	0.8986
	Shenzhen	JSRT	0.9824	0.0132	0.9461
	Montgomery & JSRT	Shenzhen	0.9727	0.0444	0.9178
4	Shenzhen	Shenzhen	0.9719	0.0454	0.9163
	Shenzhen	JSRT	0.9663	0.0536	0.9175
	Montgomery & JSRT	Shenzhen	0.9742	0.0711	0.9116
8	Shenzhen	Shenzhen	0.9612	0.0663	0.9132
	Shenzhen	JSRT	0.9364	0.0653	0.8947
	Montgomery & JSRT	Shenzhen	0.9641	0.0753	0.8816

Table 5. Performance of *Hydra* head and existing methods for CXR lung segmentation on JSRT dataset. Bold numerical values denote superior performance.

Method	IoU	Computation Time (per image)
Anatomical Atlas [40]	0.9500	20–25 s
Efficient U-Net [41]	0.9200	0.08 s
Hydra Head (Scale 2)	**0.9742**	**0.03 s**

4.3. Training Parameters and Time

Table 6 presents quantitative measurements, in terms of parameters and time, for the *Hydra*, the state-of-the-art VDSR, and the traditional baseline approach for visual task analysis. We trained all models using a Windows system with the following configuration: (1) Intel Xeon CPU E3-1275 v6 3.80 GHz processor and (2) NVIDIA GeForce GTX 1050 Ti. Keras DL framework with Tensorflow was used for model training and evaluation.

As shown in Table 6, the training parameters and time for the state-of-the-art VDSR are higher than the proposed *Hydra* trunk. Similarly, the computation time, per image, of VDSR is higher than *Hydra* trunk. In addition to the SISR models, Table 6 presents the training time and parameters as well as the computation time (per image) for *Hydra* heads and the traditional baseline approach. As can be seen in Table 6, the training parameters and time for *Hydra* classification head are notably lower than the baseline model. In case of segmentation, both *Hydra* head and the baseline segmentation model have the same number of training parameters. However, the training time of *Hydra* segmentation head, which was initialized using the trunk weights, is much lower than the baseline model. This suggests that using the weights of the *Hydra* trunk leads to faster convergence. Finally, it is important to note that although the total training time for *Hydra* (trunk and heads) is higher than the baseline approach, the majority of *Hydra* time is coming from the trunk (SISR model). However, this trunk is trained only once and can be used to simultaneously solve N visual tasks in medical images.

In addition to the traditional baseline models, we believe *Hydra* has significantly lower computational time and parameters as compared to existing methods in the literature. Specifically, given that (1) the number of parameters for most SISR models in the literature ranges from 57K to 43M [27] and (2) the number of parameters for the standalone state-of-the-art classification models ranges from 7M to 146M [42], we believe the proposed *Hydra* significantly decreases the training time and parameters for super-resolution image enhancement and visual task analysis in medical images. Besides being able to significantly reduce training and speed-up computations, similar to [43] *Hydra* can tackle the issue of the high dependence on a relatively large corpus of labelled data as follows. The *Hydra* trunk can be created using an unsupervised learning (no manual labeling) or residual learning (SISR), and then used as a shared backbone to simultaneously learn task-specific layers with relatively small numbers of training parameters; i.e., freezing the weights of the trunk and learning shallow task-specific layers can lead to enhanced performance in visual task analysis, especially for tasks with limited amounts of labeled data.

Our experimental results are promising and prove the efficiency and superiority of the proposed approach for efficient medical image analysis.

Table 6. Summary of training parameters and time for *Hydra*, VDSR, and the traditional approach.

Approach	Task	Train Parameters	Train Time (Seconds)	Test Time (Seconds)
VDSR [21]	SISR	668,227	72,490.26	0.23
Hydra	SISR (Trunk)	395,717	21,917.64	0.09
	Classification Head	295,715	180.75	0.01
	Segmentation Head	786,497	387.87	0.03
Traditional	Baseline Classification	1,573,506	650.72	0.05
	Baseline Segmentation	786,497	8238.98	0.02

5. Conclusion and Possible Extensions

HR images with fine structural details are preferred for visual task analysis in medical images. Several deep learning networks have been proposed in the literature to enhance the resolution of medical images. However, these networks are very deep and require a massive number of parameters,

which restrict them to platforms with large memory banks and computing capability. To address this problem, we propose an efficient approach, called the *Hydra*, to accelerate SISR and visual task analysis in medical images. The *Hydra* consists of a SISR trunk and task-specific heads. The trunk is a small SISR model that learns the mapping from LR to HR. The experimental results demonstrated that the performance of the *Hydra* trunk is comparable to the state-of-the-art VDSR model. After constructing the trunk, task-specific heads are appended to learn task-specific features for the key visual tasks in medical images. The proposed *Hydra* was evaluated on CXR datasets to perform image enhancement, abnormality classification, and lung segmentation. The experimental results demonstrated that the proposed approach outperformed baseline approach by a large margin. In addition, quantitative measurements of the *Hydra* showed a remarkable reduction in terms of resources and computations as compared to the baseline traditional approach and methods in the literature.

Although we only demonstrated the performance of our approach in constructing HR CXR images followed by classification and segmentation, the *Hydra* can be extended in several ways. For example, instead of using a single LR image, multiple LR images can be used as input to the SISR model (*Hydra* trunk) followed by fusing the reconstructed images as proposed in [44,45] to generate a fused HR image. Besides, the proposed *Hydra* can be extended to perform pre-processing operations in 3D images using 3D CNNs (e.g., 3D SISR [46]) as a shared trunk followed by appending the trained trunk with multiple heads for different 3D tasks. The *Hydra* can also be extended and used for other medical image applications. For example, a trunk can be trained to construct HR CT images followed by appending the trunk with task-specific layers to perform different CT tasks. In another imaging modality, a *Hydra* trunk can be trained to de-noise MRI images followed by simultaneous visual task analyses including the detection of a brain tumor, its location, and the segmentation of the tumor from the background.

Author Contributions: G.Z., S.R., and S.A. designed the framework, G.Z. and S.R. conceived the experiment(s), G.Z. and S.R. conducted the experiment(s), G.Z., S.R., and S.A. analyzed the results. G.Z. wrote the manuscript. All authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arai, K.; Kapoor, S. *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC)*; Springer: New York, NY, USA 2019; Volume 2.
2. Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. *arXiv* **2019**, arXiv:1902.06068.
3. Giger, M.L. Machine learning in medical imaging. *J. Am. Coll. Radiol.* **2018**, *15*, 512–520.
4. Shi, L.; Tashiro, S. Estimation of the effects of medical diagnostic radiation exposure based on DNA damage. *J. Radiat. Res.* **2018**, *59*, ii121–ii129.
5. Li, X.; Fu, W. Regularized super-resolution restoration algorithm for single medical image based on fuzzy similarity fusion. *EURASIP J. Image Video Process.* **2019**, *2019*, 83.
6. Xu, L.; Zeng, X.; Huang, Z.; Li, W.; Zhang, H. Low-dose chest X-ray image super-resolution using generative adversarial nets with spectral normalization. *Biomed. Signal Process. Control* **2020**, *55*, 101600.
7. Yun, S.J.; Ryu, C.W.; Choi, N.Y.; Kim, H.C.; Oh, J.Y.; Yang, D.M. Comparison of low-and standard-dose CT for the diagnosis of acute appendicitis: a meta-analysis. *Am. J. Roentgenol.* **2017**, *208*, W198–W207.
8. Mouton, A.; Breckon, T.P. On the relevance of denoising and artefact reduction in 3d segmentation and classification within complex computed tomography imagery. *J. X-ray Sci. Technol.* **2019**, *27*, 51–72.
9. Sippola, S.; Virtanen, J.; Tammilehto, V.; Grönroos, J.; Hurme, S.; Niiniviita, H.; Lietzen, E.; Salminen, P. The accuracy of low-dose computed tomography protocol in patients with suspected acute appendicitis: The OPTICAP study. *Ann. Surg.* **2020**, *271*, 332–338.

10. Rundo, L.; Tangherloni, A.; Cazzaniga, P.; Nobile, M.S.; Russo, G.; Gilardi, M.C.; Vitabile, S.; Mauri, G.; Besozzi, D.; Militello, C. A novel framework for MR image segmentation and quantification by using MedGA. *Comput. Methods Programs Biomed.* **2019**, *176*, 159–172.

11. Rundo, L.; Tangherloni, A.; Nobile, M.S.; Militello, C.; Besozzi, D.; Mauri, G.; Cazzaniga, P. MedGA: A novel evolutionary method for image enhancement in medical imaging systems. *Expert Syst. Appl.* **2019**, *119*, 387–399.

12. Lukin, A.; Krylov, A.S.; Nasonov, A. Image interpolation by super-resolution. *Proc. GraphiCon* **2006**, *2006*, 239–242.

13. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust.* **1981**, *29*, 1153–1160.

14. Li, X.; Orchard, M.T. New edge-directed interpolation. *IEEE Trans. Image Process.* **2001**, *10*, 1521–1527.

15. Sun, J.; Xu, Z.; Shum, H.Y. Image super-resolution using gradient profile prior. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 24–26 June 2008; pp. 1–8.

16. Schulter, S.; Leistner, C.; Bischof, H. Fast and accurate image upscaling with super-resolution forests. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3791–3799.

17. Li, H.; Lam, K.M.; Wang, M. Image super-resolution via feature-augmented random forest. *Signal Process. Image Commun.* **2019**, *72*, 25–34.

18. Gu, P.; Zheng, L. Fast low-dose Computed Tomography image Super-Resolution Reconstruction via Sparse coding and Random Forests. In Proceedings of the 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 24–26 May 2019; pp. 1400–1403.

19. Freeman, W.T.; Jones, T.R.; Pasztor, E.C. Example-based super-resolution. *IEEE Comput. Graph. Appl.* **2002**, *22*, 56–65.

20. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 391–407.

21. Kim, J.; Kwon Lee, J.; Mu Lee, K. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.

22. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

23. Mahapatra, D.; Bozorgtabar, B. Progressive Generative Adversarial Networks for Medical Image Super resolution. *arXiv* **2019**, arXiv:1902.02144.

24. Goodfellow, I. NIPS 2016 tutorial: Generative adversarial networks. *arXiv* **2016**, arXiv:1701.00160.

25. You, C.; Li, G.; Zhang, Y.; Zhang, X.; Shan, H.; Li, M.; Ju, S.; Zhao, Z.; Zhang, Z.; Cong, W.; et al. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Trans. Med. Imaging* **2019**, *39*, 188–203.

26. Joshi, S.H.; Marquina, A.; Osher, S.J.; Dinov, I.; Darrell, J. *Image Resolution Enhancement and Its Applications to Medical Image Processing*; Laboratory of Neuroimaging University of of California: Los Angeles, CA, USA, 2008.

27. Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H.; Liao, Q. Deep learning for single image super-resolution: A brief review. *IEEE Trans. Multimed.* **2019**, *21*, 3106–3121.

28. Gulati, T.; Sengupta, S.; Lakshminarayanan, V. Application of an enhanced deep super-resolution network in retinal image analysis. In Proceedings of the Ophthalmic Technologies XXX, San Francisco, CA, USA, 1–6 February 2020; Volume 11218, p. 112181K.

29. Christodoulidis, A.; Hurtut, T.; Tahar, H.B.; Cheriet, F. A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images. *Comput. Med. Imaging Graph.* **2016**, *52*, 28–43.

30. Li, X.; Shen, L.; Xie, X.; Huang, S.; Xie, Z.; Hong, X.; Yu, J. Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection. *Artif. Intell. Med.* **2019**, *103*, 101744.

31. Shih, G.; Wu, C.C.; Halabi, S.S.; Kohli, M.D.; Prevedello, L.M.; Cook, T.S.; Sharma, A.; Amorosa, J.K.; Arteaga, V.; Galperin-Aizenberg, M.; et al. Augmenting the National Institutes of Health chest radiograph dataset with expert annotations of possible pneumonia. *Radiol. Artif. Intell.* **2019**, *1*, e180041.

32. Jaeger, S.; Candemir, S.; Antani, S.; Wáng, Y.X.J.; Lu, P.X.; Thoma, G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant. Imaging. Med. Surg.* **2014**, *4*, 475.

33. Shiraishi, J.; Katsuragawa, S.; Ikezoe, J.; Matsumoto, T.; Kobayashi, T.; Komatsu, K.I.; Matsui, M.; Fujita, H.; Kodera, Y.; Doi, K. Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *AJR Am. J. Roentgenol.* **2000**, *174*, 71–74.

34. SCR Reference Lung Boundaries. Available online: https://www.isi.uu.nl/Research/Databases/SCR/ (accessed on 24 May 2012).

35. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402.

36. Brunet, D.; Vrscay, E.R.; Wang, Z. On the mathematical properties of the structural similarity index. *IEEE Trans. Image Process.* **2011**, *21*, 1488–1499.

37. Chollet, F. Keras. Available online: https://github.com/fchollet/keras (accessed on 19 June 2015).

38. Sharma, N.; Ray, A.K.; Sharma, S.; Shukla, K.; Pradhan, S.; Aggarwal, L.M. Segmentation and classification of medical images using texture-primitive features: Application of BAM-type artificial neural network. *Med. Phys.* **2008**, *33*, 119.

39. Han, C.; Kitamura, Y.; Kudo, A.; Ichinose, A.; Rundo, L.; Furukawa, Y.; Umemoto, K.; Li, Y.; Nakayama, H. Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection. In Proceedings of the 2019 International Conference on 3D Vision (3DV), Quebec City, QC, Canada, 16–19 September 2019; pp. 729–737.

40. Candemir, S.; Jaeger, S.; Palaniappan, K.; Musco, J.P.; Singh, R.K.; Xue, Z.; Karargyris, A.; Antani, S.; Thoma, G.; McDonald, C.J. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Trans. Med. Imaging* **2013**, *33*, 577–590.

41. Narayanan, B.N.; Hardie, R.C. A Computationally Efficient U-Net Architecture for Lung Segmentation in Chest Radiographs. In Proceedings of the 2019 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 July 2019; pp. 279–284.

42. Sultana, F.; Sufian, A.; Dutta, P. Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey. *arXiv* **2020**, arXiv:2001.04074.

43. Aviles-Rivero, A.I.; Papadakis, N.; Li, R.; Sellars, P.; Fan, Q.; Tan, R.T.; Schönlieb, C.B. GraphXNet: Chest X-Ray Classification Under Extreme Minimal Supervision. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; Springer: New York, NY, USA : 2019, pp. 504–512.

44. Li, J.; Yuan, G.; Fan, H. Multifocus image fusion using wavelet-domain-based deep CNN. *Comput. Intell. Neurosci.* **2019**, *2019*, 4179397.

45. Aymaz, S.; Köse, C.; Aymaz, Ş. Multi-focus image fusion for different datasets with super-resolution using gradient-based new fusion rule. *Multimed. Tools Appl.* **2020**, *79*, 13311–13350.

46. Georgescu, M.I.; Ionescu, R.T.; Verga, N. Convolutional Neural Networks with Intermediate Loss for 3D Super-Resolution of CT and MRI Scans. *arXiv* **2020**, arXiv:2001.01330.