CNN-based morphological decomposition of X-ray images for details and defects contrast enhancement

Tahani Madmad, Nicolas Delinte, Christophe De Vleeschouwer
UCLouvain, ICTEAM, Belgium *
{tahani.madmad, nicolas.delinte, christophe.devleeschouwer}@uclouvain.be

Abstract

This paper introduces a new learning based framework for X-ray images that relies on a morphological decomposition of the signal into two main components, separating images into local textures and piecewise smooth (cartoon) parts. The piecewise smooth component corresponds to the spatial variation of the average density of the objects, whereas the local texture component presents the inspected objects singularities. Our method builds on two convolutional neural network (CNN) branches to decompose an input image into its two morphological components. This CNN is trained with synthetic data, generated by randomly picking piecewise smooth and singular patterns in a parametric dictionary and enforcing the sum of the CNN branches to approximate the identity mapping. We demonstrate the relevance of the decomposition by enhancing the local textures component compared to the piecewise smooth one. Those enhanced images compare favorably to the ones obtained with existing works destined to visualize High Dynamic Range (HDR) images such as tone-mapping algorithms.

1. Introduction

High Dynamic Range (HDR) imaging offers the ability to capture weakly contrasted details, due to a relatively fine quantization of the intensity range, making it especially popular and relevant in X-ray imaging. This is because the intensity level of X-ray images does reflect the X-ray absorption power of the observed object [45]. Therefore, a high dynamic range of intensities is required to simultaneously capture the large absorption discrepancies observed in a scene (between high density objects and X-ray transparent ones) and the small variations of X-ray absorption induced by the object surface texture or by small object defects such as cracks in metal. In medical and non-destructive evaluation imaging systems, X-ray images are recorded with 16 bits per pixel or more, which corresponds to 65536 possible intensity levels per pixel [27], whereas most medical displays only support 8 bits (256 levels of gray) per pixel [34]. Even though several manufacturers offer new display systems with higher bit depth [26], the visualization of images is itself limited by the human visual system capabilities: the number of gray levels that an average user can perceive as distinct [26] is indeed limited to around 1000 levels of distinguishable intensities for experts. Therefore, a monitor offering an intensity range equivalent to the dynamic range of the image is not enough to make the weakly contrasted image structures visually perceptible.

In this context, tone-mapping has been largely investigated in previous works to preserve the image details when mapping the high dynamic range of the scene to the displayable range [9]. Many of those algorithms rely on the fact that the range of intensities in a local neighborhood is extremely restricted compared to the dynamic range of the entire HDR image. This is the main reason why they typically design different mapping functions for distinct regions of the image. However, adapting different mappings for neighboring regions, potentially leads to aliasing effects. Moreover, in X-ray imaging the intensity is related to the physical characteristics of the object and more precisely to its density [45] as described by the equation 1. For this reason, adopting different gray levels to represent the same material density in distinct parts of the object is not acceptable. To address this problem, we assume that X-ray images consists of the sum of two components. The first one is denoted background component, and corresponds to the object's average power of absorption. The second component reflects the fine local variations of the absorption power induced by the object singularities, e.g. including the industrial defects (fractures, inclusions, shrinkage etc.) and the medical abnormalities (nodules, fractures, variant anatomic abnormalities etc). Our original approach is based on the design of a two-branch CNN architecture to predict these two components from an input X-ray image. This decomposition of X-ray images is then exploited to improve the visibility of

*Part of this work has been funded by the 'Pôle Mecatech ADRIC' Walloon Region project, and by the Belgian NSF

the singularities and the weakly contrasted details by amplifying their corresponding component.

To tackle the absence of ground-truth decomposition of real X-ray images, we propose to train our CNN based on a synthetic dataset that we construct by summing up the piecewise smooth backgrounds with synthetic spot and fractures. The training of our network is then based on the explicit supervision offered by synthetic data, for which the ground truth components are known by construction. Our trained model has been tested on both synthetic and real datasets even if the network was not trained on real X-ray images. Improved visualization of the image singularities has been obtained by amplifying the predicted singular image component, resulting in a contrast enhancement that compares favorably to state-of-art methods.

The rest of our paper is organized as follows. Section 2 briefly surveys the related works associated to mapping X-ray images and image morphological decomposition, in addition to CNN-based frameworks for X-ray enhancement and visualization. Section 3 provides a formal description of our proposed framework. Based on the experiments presented in the first part of section 4, we demonstrate the effectiveness and the relevance of our approach in the second part of this section. Section 5 concludes .

2. Related Works

Tone-mapping Algorithms. The question of tone-mapping algorithms has been widely studied in the recent years [9, 10, 17]. There are mainly two categories of methods developed to solve this issue: Global and local operators [9]. Global operators apply a predefined mapping function to all HDR pixel intensities. In other words, equal intensity values in the input image are assigned to the exact same level of the dynamic range in the output image [54, 50, 7, 39, 11]. Even though this kind of methods is easy to implement and generally preserves the intensity relative ordering of the original pixel intensities, the local contrast of fine variations remains insufficient and the resulting visual attenuation of the details in the medical field or in industrial non-destructive evaluation can have major drawbacks in each field [51, 25]. In contrast, local operators process pixel intensities as a function of their local context. Even though local methods better preserve weakly contrasted details, they usually suffer from higher computational cost, and poor global consistency of the resulting image, inducing eventual visual artifacts [9].

In the domain of X-ray imaging, mapping the measured intensities to the displayable ones is a highly ill-posed problem. Mathematically, the only known is the X-ray original image thus there are two unknowns that corresponds to the transformation function and the ideal output image. In order to tackle this problem, conventional frameworks destined to visualize X-ray images rely in practice on physical priors linked to the shape and the nature of the object or the anomalies as extra constraints [31]. By definition, those methods are destined to process a well defined type of image content (e.g chest X-ray) and can not be considered to process a wild variety of X-ray images. The bilateral histogram equalization introduced by [32] offers a good balance between global and local tone-mapping by adapting the mapping locally based on the global distribution of intensities but yet it is time consuming and can introduce some ringing artefacts when it comes to process filiform structures. In our study, we exploit the deep neutral networks to indirectly learn such transformation without defining a formal model that describes those variations.

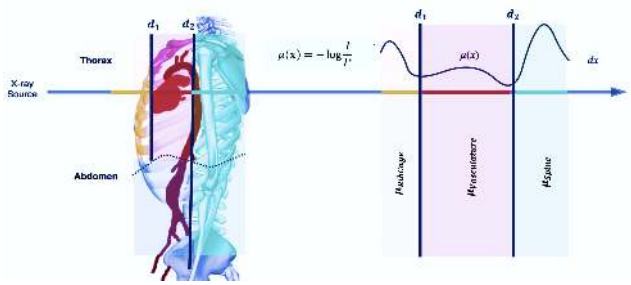

Figure 1: **X-ray image formation process illustrated based on a synthetic CT volume of the abdomen. [Copied from [1]]**

Image Decomposition. The majority of local algorithms that aim to fit an HDR image into a low dynamic range screen are based on a layer decomposition of the image [28, 2, 8, 43]. These approaches aim to enhance X-ray image visualization based on a decomposition of the image into two main layers, where the first layer is obtained by filtering out the fine structures in the HDR image. The second layer corresponds to the residues, obtained by subtracting the filtered image from the original HDR image. The final output is then reconstructed by scaling down the first layer so as to fit the display dynamic range, while preserving the details by adding the residue to the scaled image. Several variants of this approach have been proposed [61, 60, 56, 13] using different kinds of filters [38, 55, 42, 23]. Unfortunately, these signal processing methods based on filtering have to be tuned depending on the type of defect to be extracted. In practice, the filtering based methods that aim to decompose the image into two layers are not only parameter sensitive, but the choice of the filter is itself intrinsically linked to the class of signal to be extracted. In contrast, our approach targets universality, and is inspired by the image morphological decomposition frameworks [31]. One of the pioneering works in this field is the morphological component analysis (MCA) based on sparse representation of signals introduced by [47, 46]. The MCA is based on the assumption that each signal is the

linear mixture of several layers, that are morphologically distinct. The core assumption of this method is that every component behavior emanates from a sparse representation construction of a dictionary of atoms [31]. In this context, the separation of an image into its cartoon (background) and texture layers have been popular among the community of computer vision and machine learning and several approaches have been proposed to learn the morphological decomposition images [47, 41, 21, 24, 57]. In this context, [30] proposed an intrinsic image decomposition based on a separation of the natural input image into a reflecting and a shading component for image-to-image transformations. Further, [21] addresses the task of single image denoising based on a semantic decomposition of the image. In [22] the authors suggest a single image rain removal method based on image decomposition by learning dictionaries for the high-frequency components. Recently, [59] reformulate the problem of image dehazing as a restoration of the image base component. Based on the learning of the decomposition of a hazy image into the base and the detail components, the authors achieved haze removal (or dehazing). Building on these ingredients, we will provide leaning based framework for X-ray image local contrast enhancement and tone-mapping via the X-ray Morphological Decomposition (XMD).

Convolutional Neural Networks. A major factor that affects the quality of the image according to quality measurement literature is the change of spatial correlation [52]. Research in deep learning has shown that the hidden layers of a convolutional neural network can capture a variety of spatial correlation properties of the input image and describe its fine variations and local structure [20]. As one of the most important goals of X-ray visualization, the preservation of the integrity and the overall appearance of the image is necessary. It follows that the spatial correlations of the reconstructed output should be consistent with those of the input [19]. As the deep features of a CNN capture the spatial correlations of the input image, our research on the design of a learning based operator for local structures enhancement will be based on Convolutional Neural Networks. In the field of image enhancement, CNNs have been exploited in several applications because of their ability to make a decision in function of the local structure analysis of the image. A key advantage of CNNs consists of detecting local features at different positions in the input feature maps with learnable kernels [58]. As a matter of fact, CNNs have proven their efficiency for multiple HDR imaging applications such as reconstructing HDR from a single-exposure Low Dynamic Range (LDR) [35] or HDR reconstruction from several low exposure images [12]. CNNs have also been widely used to learn image-to-image transformations that enhance the visual quality of images e.g tone-mapping, denoising and colorization [19, 14, 15, 5]. Only few mapping operators destined for X-ray images have been proposed in the literature [25, 32, 21]. This is mainly due to the ground truth data unavailability [16]. In the existing frameworks, this issue is typically tackled considering a pre-defined tone-mapping operator as an ideal transformation to generate the ground truth. Unfortunately this training strategy is characterized by a strong restriction of the performances and the domain of application. Unlike the usual tasks that aim to offer a high level vision, the mapping of X-ray images is mainly physics based [18]. Consequently, the design of a learning based method that aims to better map the X-ray intensities to the available intensities should heavily rely on physics based assumptions linked to the formation process of X-ray images. For this reasons, and in the aim of designing a generic method destined to X-ray images visualization, we do not consider a framework that requires an explicit prior on the content of the X-ray image. Instead, we adopt a generic learning-based strategy based on a morphological decomposition of the single input X-ray image. In addition, we suggest to handle the lack of reference images by relying on a synthetic dataset to learn how to decompose the input into physics based components.

3. The X-ray Morphological Decomposition (XMD)

This section first refers to the X-ray formation process to motivate our proposed decomposition of the image into two components. It then presents our envisioned CNN architecture for X-ray image decomposition, and explains how this network can be trained based on synthetic data to promote the decomposition of its input into morphological components.

3.1. X-ray Image Formation Process

As illustrated in Figure 1, the X-ray intensity measured in a pixel p can be formally written as:

$$I(p) = I^*(p) \cdot \exp(-\int_{\mathcal{T}} \mu(x)dx), [49] \qquad (1)$$

where $I^*(p)$ denotes the X-ray intensity measured by the receptor at the pixel p in absence of any interaction/attenuation with the materials, and $\mathcal{T}(p)$ the path between the source and the pixel p, denoted $\mathcal{T}$ to simplify the notations. The path is indexed by x, so that $\mu(x)$ defines the X-ray energy absorption coefficient in position x along the trajectory $\mathcal{T}$.

Considering that we are interested in differentiating the X-ray attenuation induced by the core object from the one caused by the object singularity/texture, let $A^o(p) = \exp(-\int_{\mathcal{T}} \mu^o(x)dx)$ denote the absorption induced by the core and 'ideal' (without singularities) object and and $A^s(p) = \exp(-\int_{\mathcal{T}} \mu^s(x)dx)$ the singularities. In these

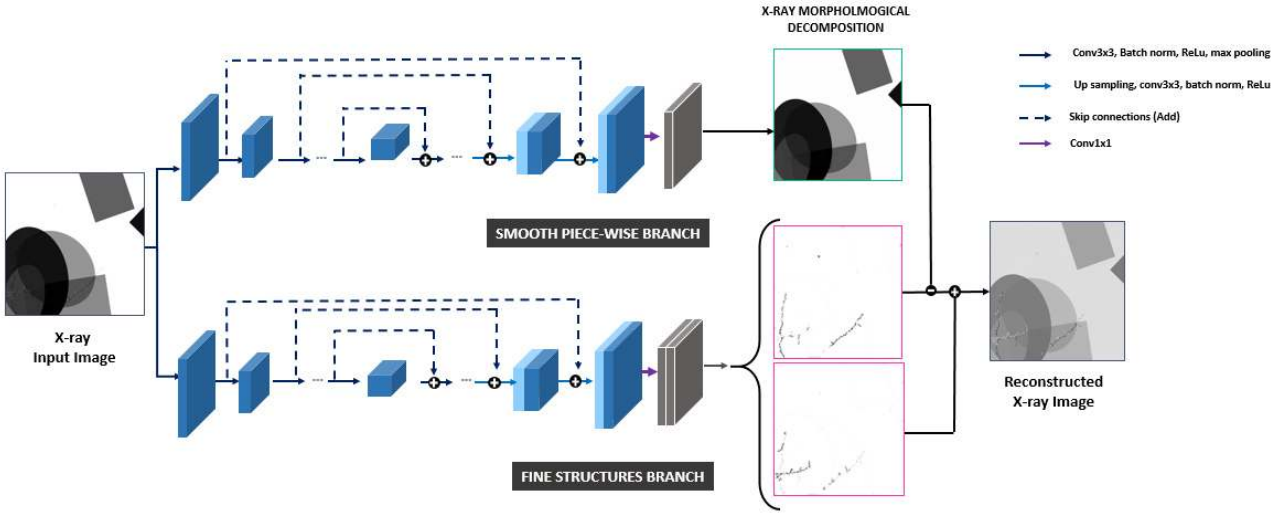

Figure 2: Overview of the network architecture. The network is composed of two main branches, one for the prediction of the piecewise component and the other for fine structures, both of which are based on a 6-layer deep U-Net with skip connections. The piecewise branch has one output channel whereas the fine structures branch has two output channels.

notations, $\mu^o(x)$ denotes the absorption coefficient associated to an object without fine texture nor singularities, while $\mu^s(x) \triangleq \mu(x) - \mu^o(x)$ reflects the attenuation increase or decrease induced by singularities. In most positions along the path, $\mu^s(x) = 0$, so that $A^s(p) \simeq 1$.

To reflect the fact that singularities only slightly affect the X-ray absorption, we write:

$$A^s(p) = 1 + \epsilon(p), \tag{2}$$

so as to reflect that the singularity slightly increases ($\epsilon^s(p)$ is a small negative value) or decreases ($\epsilon^s(p)$ is a small positive value) the absorption induced by the object without singularity. Hence, we have:

$$I(p) = I^*(p) \cdot A^o(p) + I^*(p) \cdot A^o(p) \cdot \epsilon(p), \tag{3}$$

and the X-ray image can be envisioned as the sum of two components, respectively corresponding to the (piecewise smooth) object and to its texture or singularity. This decomposition makes it trivial to improve the visibility of the singularities, simply by multiplying the corresponding component by a constant factor. Our work aims at deriving a network that decomposes an input into those two components.

3.2. The CNN Architecture for Image Decomposition

The network used to decompose an input X-ray image into its components is based upon the U-Net architecture, which was originally developed for the segmentation of biomedical images. The U-Net architecture has also been considered as a generative model, e.g. for denoising purposes [4]. Our proposed model is presented in Figure 2. It is composed of two branches that operate in parallel. The first branch aims at reconstructing the piecewise smooth component, whereas the second branch is responsible of recovering the fine structures magnitude components, induced by textures and singularities. Each branch consists in a U-Net, whose convolutional blocks at each resolution are composed of six operations: a 2D convolution, a batch normalization, a rectified linear unit (ReLu) [36], followed by a second convolution, batch normalization and ReLu. The neural network involves six resolution levels to ensure the coverage of a large receptive field. As illustrated in Figure 2, the output of the second branch consists in two images that respectively define the positive and negative values of the fine structures variation component. Since the components are expected to decompose the input, their sum should reconstruct the input image, which provides a natural constraint when defining the losses involved in the training of the network, as detailed in the next section.

3.3. Morphological Components Prediction

Assumptions and Hypothesis. The training of our CNN builds on two main assumptions. (i) Based on the X-ray image formation process defined by Equation 3, we assume that an X-ray HDR image is composed of two main components. The first one is defined as a piecewise smooth class of signals representing the object's average power of absorption in X-ray images (Background/cartoon). The second one corresponding to the fine structures including fabric

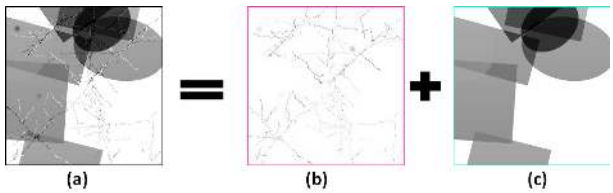

Figure 3: Synthetic X-ray image generation. (a) X-ray synthetic image. (b) Fine structure variations component. (c) Smooth piecewise component.

defects, fractures and other meaningful singularities of the signal (Textures). (ii) Based on the physical characteristics of the X-ray signal, we assume that each component of an X-ray HDR image can be described by a linear combination of functions with specific shapes, forming a dictionary [41]. Following the terminology adopted for blind source separation, the components are then named morphological components since their morphological characteristics are defined by the shapes of the dictionary functions.

Generating the Synthetic Dataset. We propose to create a synthetic dataset mimicking the desired morphological decomposition of X-ray images. As a consequence of the X-ray image model presented in Equation 3, each image $x(u, v)$ in this dataset is defined as the sum of two components $x_1(u, v)$ and $x_2(u, v)$ that are synthesized by combining functions picked-up in dictionaries respectively chosen to reproduce the piecewise smooth nature of X-ray absorption maps, and to reflect the singularities encountered in real images (defects, cracks, nodules).

Formally, letting $\mathcal{D}$ denote a dictionary of smooth functions with finite spatial support, we have:

$$x(u, v) = x_1(u, v) + x_2(u, v), \tag{4}$$

with

$$x_1(u, v) = \sum_k a_k \cdot d_k(u, v), d_k \in \mathcal{D}, \tag{5}$$

and

$$x_2(u, v) = x_1(u, v) \cdot \epsilon(u, v), \tag{6}$$

where $\epsilon(u, v)$ corresponds to small positive (absorption power increase) and negative (absorption power decrease) values that are spatially distributed to reflect typical X-ray image singularities.

In practice, in our experiments, the dictionary $\mathcal{D}$ has been defined based on elliptic and rectangular geometric shapes. The shapes, displayed in Figure 3, have a varying orientation, size and shade of gray, to mimic materials of different shape and density represented in the same X-ray image. In addition, an intensity gradient was added randomly to some images to simulate the possibility of a increasing absorption power, due to increased object thickness, for example.

Two types of functions have been considered to generate the $\epsilon(u, v)$ singularities (defects) : (i) Branches, as illustrated in Figure 3, mimic bone fractures, respiratory branches, shrinkage or cracks that are defects usually found in casting pieces for example. Each branch is created using a random walk approach, with all parameters being randomly selected within a pre-defined window at each step of the random walk. These parameters include the branch-off probability, the branch death probability and the percentage of intensity gain or loss at each step. (ii) Round or elongated soft shady spots are generated to mimic nodules or cystic masses for the medical field and gas holes, porosities, inclusions and dross in the industrial domain. They are also added at random positions with varying sizes and intensities. An example of ϵ image can be seen in Figure 3.

Network Training. During training, the branches of the network are updated so as to minimize the reconstruction error of each component (when available, i.e. for synthetic images only), but also the reconstruction error of their sum (which has to be equal to the input, available both for real and synthetic images). Therefore, four losses are introduced. L_1 compares the prediction of the first branch to the reference piecewise smooth component. L_{2+} and L_{2-} ensure that the two maps predicted by the second branch respectively reconstruct (in absolute value) the positive and negative parts of the second reference component. L_{rec} measures how well the input image is approximated by the sum of the two predicted components.

We adopt the Multi-scale Structural Similarity Index Measure (MS-SSIM) [53] for each loss.
The global objective function that combines the three losses is defined as:

$$L_g = \frac{\alpha \cdot (L_{2+} + L_{2-}) + \beta \cdot L_1 + \gamma \cdot L_{rec}}{2 \cdot \alpha + \beta + \gamma}, \tag{7}$$

with $\alpha, \beta, \gamma \in [0, 1]$.

4. Experiments and Results

This section validates our proposed approach both on synthetic and real images[1]. For synthetic image, since we have access to the ground-truth components, we verify that the trained model generalizes properly on unseen images. In contrast, for real images, we verify that the two components that are predicted by the network sum-up to the input image, and assess the relevance of the partition. Therefore, we reconstruct images in which the local structure component is amplified compared to the piecewise smooth one. This

[1]Python source code for generating the synthetic dataset, as well as the Pytorch implementation of the network is available on https://github.com/tahanimadmad/CNN-Based-X-ray-Morphological-Decomposition-.

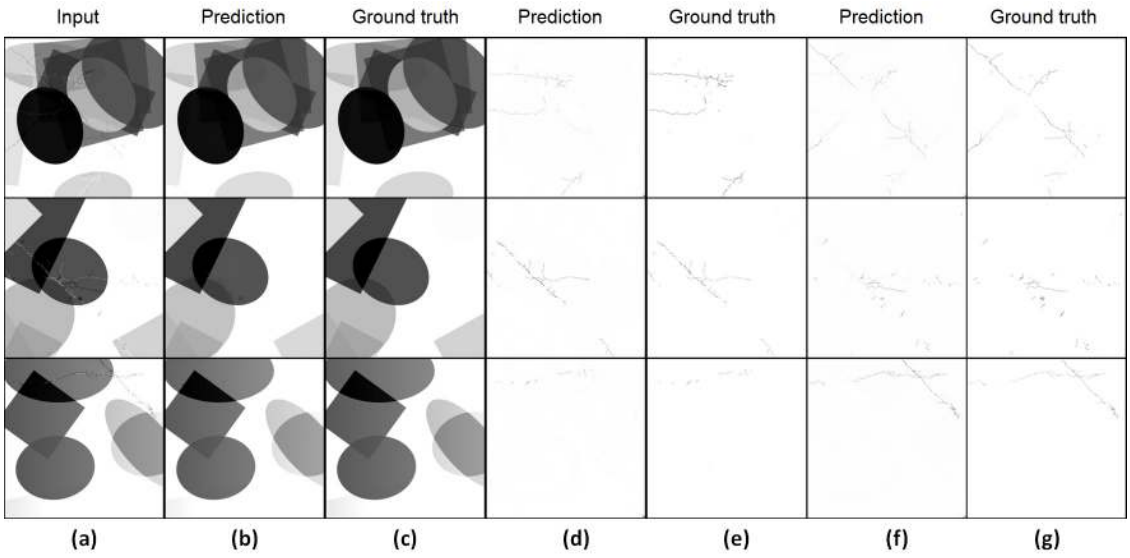

Figure 4: Results of the X-ray morphological decomposition (XMD). (a) Input. (b) The smooth piecewise prediction. (c) The smooth piecewise target. (d) The positive fine structures magnitude prediction. (e) The positive fine structures magnitude target. (f) The negative fine structures magnitude prediction. (g) The negative fine structures magnitude target.

results in a tone-mapped image, in which local structures have been magnified, and that can be assessed using conventional objective image quality assessment metrics. The images considered in the real image dataset include archaeological images, medical images and industrial images. The large application range of those images allows to demonstrate that, despite it has been trained on synthetic data, our method is not restricted to a specific kind of imaged objects.

4.1. Training Set-up and Parameters

We adopt the Rectified Adam optimizer (RAdam) [29] with the following hyper-parameters : an initial learning rate `lr`=1e-5 and β=(0.9, 0.999). The learning rate is adjusted using the a scheduler on plateau using the torch optimization package. In addition, an early stopper, interrupting the learning process when the loss has not decreased for a specified amount of epochs, has been implemented as proposed by [48] The loss functions, is based on the MS-SSIM with a window size of 5 and a with a maximum scale level of 7 ([40] implementation). To have a loss behavior similar to the MSE-like methods, we consider 1-MS SSIM instead of the MS-SSIM value. The coefficients defined in Equation 7 have been fixed as follows: α = 1, β = 2 and γ = 1. Instead of using a dropout or a parameter norm penalty to ensure the regularization, we consider a data augmentation on the fly during its training phase and we do not forget to activate the shuffle of the data in the data loader of both training and validation.The early stopper ends the learning after 125 epochs. The size of the synthetic dataset used for training = 1400, validation = 400 and test = 200, and images are sized 256x256.

4.2. Synthetic X-ray Images Decomposition

The morphological decomposition of the synthetic dataset has been validated using a sample of images that were not used in the training or validation sets, used to train the neural network. Result samples are displayed in Figure 4, where the input (a) is decomposed into three components : the smooth piecewise component prediction (b) and the positive (d) and negative (f) fines structures. The differences between the predictions (b-d-f) and their corresponding targets (c-e-g) are difficult to distinguish at a first glance, which is an indicator of the quality of the decomposition. However, upon closer inspection, there are some small variations of pixel intensities between the targets and the predictions, as well as some fine structures that have not been entirely removed from the smooth piecewise by the network. An example of these remaining fine structures is the shady spot in the bottom of the dark grey ellipse on the second row of images.

4.3. Generalization to Real X-ray Images

Despite the simplicity of the shapes and fine structures considered in the synthetic dataset used to train the network, the decomposition learned by the network generalizes surprisingly well to real X-ray images. Figure 5 presents how the network extracts morphological components in a multitude of X-ray images coming from different domains, such as medical imaging and industrial applications. Since the images are separated into three different components, it is

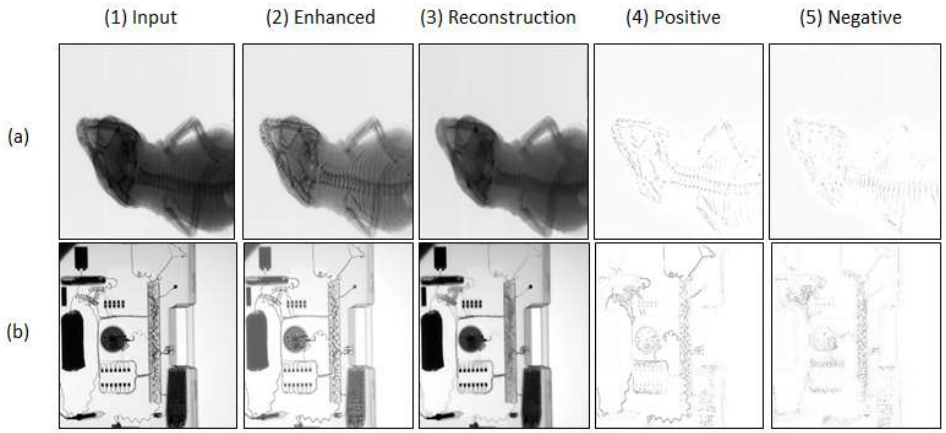

Figure 5: Results of the X-ray morphological decomposition (XMD) for a set of 2images archaeological and security applications. (a) Archaeological X-ray image. (b) Industrial X-ray image. (1) Input. (2) Our local contrast enhancement. (3) The smooth piecewise prediction. (4) The positive fine structures magnitude prediction. (5) The negative fine structures magnitude target.

possible to enhance the contrast in areas presenting fine details (more information on the contrast enhancement is presented in Section 4.4). The enhanced images are presented in Figure 5-(2). In these images, regions with an initial low contrast (such as the head and belly of the animal on row (a) and the circular object on row (b)) now present more visible fine details.

4.4. Enhancing Local Structures in X-ray Images

In this part, we leverage image decomposition to improve the visualization of details and defects in X-ray images. Therefore, we propose to highlight the local structures of X-ray images, which are supposed to represent defects or abnormalities. Those local structures are generally invisible when the image is displayed on the standard display, using a linear mapping. By reducing the intensity of the smooth piecewise component compared to the intensity of the fine details, regions that are over- or under-exposed can benefit from an enhanced contrast and an increased fine details visibility. Examples of an enhanced contrast using our algorithm can be seen in Figure 5-[2] and 6-[f].

For the image objective quality assessment, we adopted widely used and well-known measures including the Tone-Mapped Image Quality Index (TMQI) and the High Dynamic Range Visible Difference Predictor (HDR-VDP) in its second version [37]. The TMQI is a widely used metric for quality assessment of tone-mapped images, it evaluates the structural fidelity and naturalness of transformed images. the TMQI takes 1 for the best quality score and 0 for the worst score. The HDR-VDP-2 [33] is a calibrated visual metric for visibility and quality predictions in all luminance conditions, known to evaluate if the differences between the HDR input image and the reconstructed image are visible by the human eye in terms of the "Just Noticeable Differences" (JND) introduced by [6] and used in the medical field to describe how the human visual system views a medical image [3]. This metric takes the value of 10 as a maximum value and represents the best quality score. Table 1 and 2 show how mapping X-ray images from different backgrounds based on our X-ray morphological decomposition better preserves the global appearance of the image and enhances the local structures of the image.

Table 1: Tone-mapped Image Quality Index (TMQI)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Fattal [13]	0.879	0.798	0.799	0.879	0.900	0.697
Ashikhmin [2]	0.844	0.892	0.798	**0.977**	**0.961**	0.689
Reinhard [44]	0.913	0.927	0.920	0.902	0.934	0.799
Durand [8]	0.936	0.874	0.895	0.933	0.912	0.687
XMD (ours)	**0.980**	**0.932**	**0.984**	0.944	0.959	**0.856**

Table 2: HDR Visual Difference Predictor (HDR-VDP 2)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Fattal [13]	5.155	4.632	4.737	4.888	6.419	4.081
Ashikhmin [2]	7.493	6.396	6.254	7.478	7.382	6.162
Reinhard [44]	8.066	9.046	7.569	8.521	7.437	8.424
Durand [8]	6.207	4.736	5.096	6.034	5.139	5.521
XMD (ours)	**9.684**	**9.317**	**9.033**	**9.582**	**8.424**	**8.838**

Interestingly, our method offers the end-user the opportunity to evolves progressively from the visualization of the absorptive power variations rendered by a linear mapping, to the observation of the fine textural and structural object details.

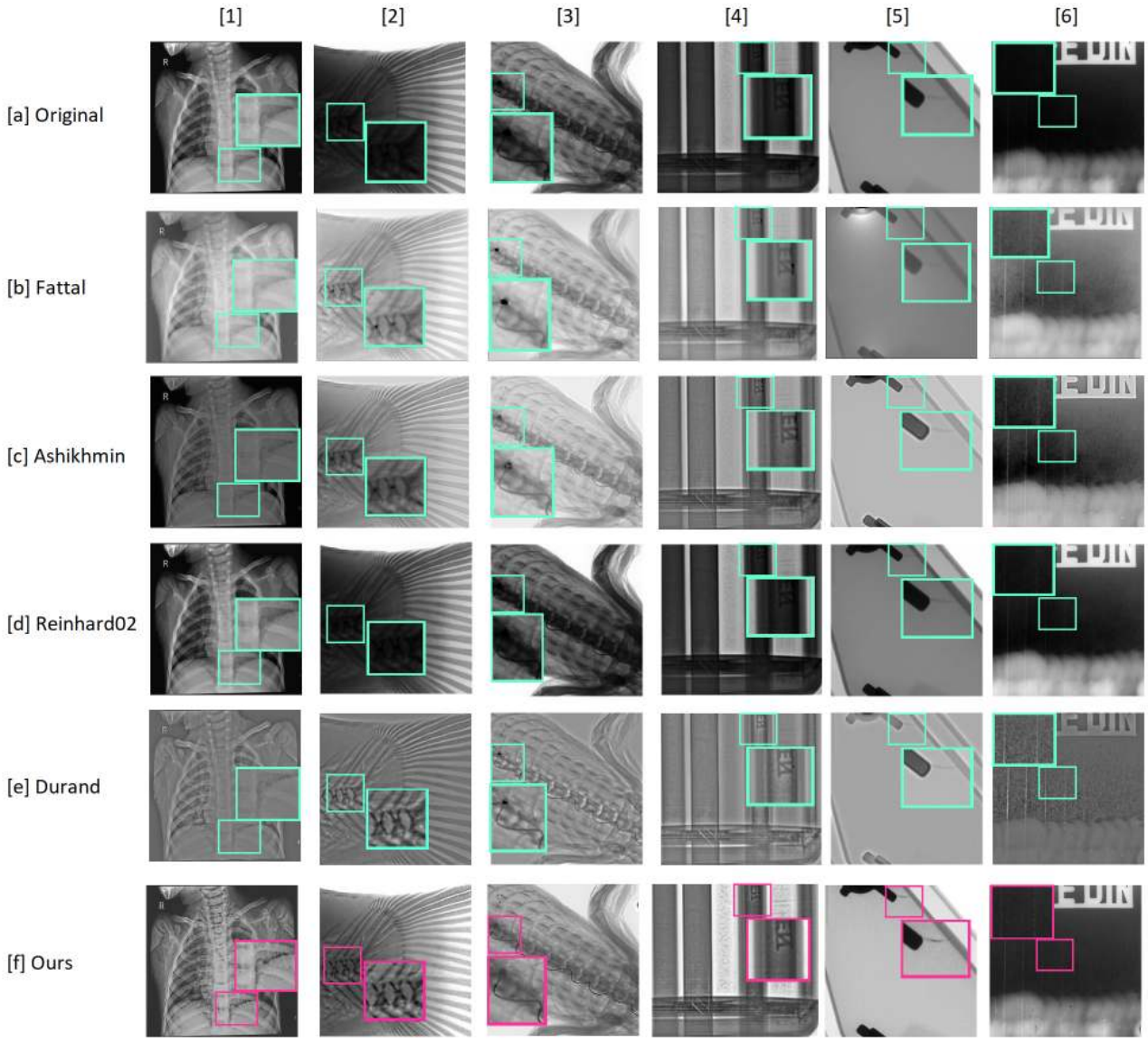

Figure 6: X-ray images tone-mapping based on our morphological decomposition learning and comparison for a set of 6 X-ray HDR images from medical, industrial and archaeological applications. (a) Original HDR image. (b) Fattal [13]. (c) Ashikhmin [2]. (d) Reinhard02 [44]. (e) Durand [8]. (f) XMD (ours) .

5. Conclusion

This paper has introduced an original CNN-based strategy to decompose X-ray images into their morphological components. Our network is trained on a synthetic dataset generated based on the X-ray image formation process to tackle the lack of ground truth. The X-ray morphological decomposition is shown to generalize well on real data. We demonstrate the relevance of the decomposition by enhancing the local textures component compared to the piecewise smooth one. Those enhanced images compare favorably to the ones obtained with existing works geared towards the visualization of HDR images such as tone-mapping algorithms.

References

[1] Shadi Albarqouni, Javad Fotouhi, and Nassir Navab. X-ray in-depth decomposition: Revealing the latent structures. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 444–452. Springer, 2017. 4322

[2] Michael Ashikhmin and Jay Goyal. A reality check for tone-mapping operators. *ACM Transactions on Applied Percep-*

tion (TAP), 3(4):399–411, 2006. 4322, 4327, 4328

[3] Peter GJ Barten. Physical model for the contrast sensitivity of the human eye. In *Human Vision, Visual Processing, and Digital Display III*, volume 1666, pages 57–72. International Society for Optics and Photonics, 1992. 4327

[4] Benoit Brummer and Christophe De Vleeschouwer. Natural image noise dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 4324

[5] Jiawen Chen, Andrew Adams, Neal Wadhwa, and Samuel W Hasinoff. Bilateral guided upsampling. *ACM Transactions on Graphics (TOG)*, 35(6):1–8, 2016. 4323

[6] Scott J Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing, and Digital Display III*, volume 1666, pages 2–15. International Society for Optics and Photonics, 1992. 4327

[7] Frédéric Drago, Karol Myszkowski, Thomas Annen, and Norishige Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, volume 22, pages 419–426. Wiley Online Library, 2003. 4322

[8] Frédo Durand and Julie Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *ACM transactions on graphics (TOG)*, volume 21, pages 257–266. ACM, 2002. 4322, 4327, 4328

[9] Gabriel Eilertsen, Rafal Konrad Mantiuk, and Jonas Unger. A comparative review of tone-mapping algorithms for high dynamic range video. 36(2):565–592, 2017. 4321, 4322

[10] Gabriel Eilertsen, Jonas Unger, and Rafal K Mantiuk. Evaluation of tone mapping operators for hdr video. In *High Dynamic Range Video*, pages 185–207. Elsevier, 2016. 4322

[11] Dragomir M El Mezeni and Lazar V Saranovac. Enhanced local tone mapping for detail preserving reproduction of high dynamic range images. *Journal of Visual Communication and Image Representation*, 53:122–133, 2018. 4322

[12] Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deep reverse tone mapping. *ACM Trans. Graph.*, 36(6):177–1, 2017. 4323

[13] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, volume 27, page 67. ACM, 2008. 4322, 4327, 4328

[14] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics (TOG)*, 35(6):1–12, 2016. 4323

[15] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)*, 36(4):1–12, 2017. 4323

[16] Eli Gibson, Wenqi Li, Carole Sudre, Lucas Fidon, Dzhoshkun I Shakir, Guotai Wang, Zach Eaton-Rosen, Robert Gray, Tom Doel, Yipeng Hu, et al. Niftynet: a deep-learning platform for medical imaging. *Computer methods and programs in biomedicine*, 158:113–122, 2018. 4323

[17] Philippe Hanhart, Marco V Bernardo, Manuela Pereira, António MG Pinheiro, and Touradj Ebrahimi. Benchmarking of objective quality metrics for hdr image quality assessment. *EURASIP Journal on Image and Video Processing*, 2015(1):39, 2015. 4322

[18] Randolf Hanke, Theobald Fuchs, and Norman Uhlmann. X-ray based methods for non-destructive testing and material characterization. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 591(1):14–18, 2008. 4323

[19] Xianxu Hou, Jiang Duan, and Guoping Qiu. Deep feature consistent deep image transformations: Downscaling, decolorization and hdr tone mapping. *arXiv preprint arXiv:1707.09482*, 2017. 4323

[20] Xianxu Hou, Linlin Shen, Ke Sun, and Guoping Qiu. Deep feature consistent variational autoencoder. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1133–1141. IEEE, 2017. 4323

[21] D. Huang, L. Kang, Y. F. Wang, and C. Lin. Self-learning based image decomposition with applications to single image denoising. *IEEE Transactions on Multimedia*, 16(1):83–93, 2014. 4323

[22] D. Huang, L. Kang, M. Yang, C. Lin, and Y. F. Wang. Context-aware single image rain removal. In *2012 IEEE International Conference on Multimedia and Expo*, pages 164–169, 2012. 4323

[23] H Gökhan İlk, Onur Jane, and Özlem İlk. The effect of laplacian filter in adaptive unsharp masking for infrared image enhancement. *Infrared Physics & Technology*, 54(5):427–438, 2011. 4322

[24] Li-Wei Kang, Chia-Hung Yeh, Duan-Yu Chen, and Chia-Tsung Lin. Self-learning-based signal decomposition for multimedia applications: A review and comparative study. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, pages 1–9. IEEE, 2014. 4323

[25] Hyo-Young Kim, Seung Park, Yong-Goo Shin, Seung-Won Jung, and Sung-Jea Ko. Detail restoration and tone mapping networks for x-ray security inspection. *IEEE Access*, 8:197473–197483, 2020. 4322, 4323

[26] Tom Kimpe and Tom Tuytschaever. Increasing the number of gray shades in medical display systems—how much is enough? *Journal of digital imaging*, 20(4):422–432, 2007. 4321

[27] Michele Larobina and Loredana Murino. Medical image file formats. *Journal of digital imaging*, 27(2):200–206, 2014. 4321

[28] Zhetong Liang, Jun Xu, David Zhang, Zisheng Cao, and Lei Zhang. A hybrid l1-l0 layer decomposition model for tone mapping. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4758–4766, 2018. 4322

[29] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*, 2019. 4326

[30] Yunfei Liu, Yu Li, Shaodi You, and Feng Lu. Unsupervised learning for intrinsic image decomposition from a single image. In *Proceedings of the IEEE/CVF Conference*

on Computer Vision and Pattern Recognition, pages 3248–3257, 2020. 4323

[31] Fadili M. J., Starck J, Bobin J., and Moudden Y. Image decomposition and separation using sparse representations: An overview. *Proceedings of the IEEE*, 98(6):983–994, 2010. 4322, 4323

[32] Tahani Madmad and Christophe De Vleeschouwer. Bilateral histogram equalization for x-ray image tone mapping. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 3507–3511. IEEE, 2019. 4322, 4323

[33] Rafał Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)*, 30(4):1–14, 2011. 4327

[34] Rafał K Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. High dynamic range imaging. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–42, 1999. 4321

[35] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In *Computer Graphics Forum*, volume 37, pages 37–49. Wiley Online Library, 2018. 4323

[36] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010. 4324

[37] Manish Narwaria, Rafal Mantiuk, Mattheiu P Da Silva, and Patrick Le Callet. Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501, 2015. 4327

[38] Mariya Nazarkevych, Ivanna Klyujnyk, Iryna Maslanych, Bohdana Havrysh, and Hanna Nazarkevych. Image filtration using the ateb-gabor filter in the biometric security systems. In *2018 XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, pages 276–279. IEEE, 2018. 4322

[39] Sumanta N Pattanaik, James A Ferwerda, Mark D Fairchild, and Donald P Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 287–298. ACM, 1998. 4322

[40] Jorge Pessoa. Pytorch-msssim. https://github.com/jorge-pessoa/pytorch-msssim.git, 2017. 4326

[41] Gabriel Peyré, Jalal Fadili, and Jean-Luc Starck. Learning the morphological diversity. *SIAM Journal on Imaging Sciences*, 3(3):646–669, 2010. 4323, 4325

[42] Andrea Polesel, Giovanni Ramponi, and V John Mathews. Image enhancement via adaptive unsharp masking. *IEEE transactions on image processing*, 9(3):505–510, 2000. 4322

[43] Erik Reinhard and Kate Devlin. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization & Computer Graphics*, (1):13–24, 2005. 4322

[44] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM transactions on graphics (TOG)*, 21(3):267–276, 2002. 4327, 4328

[45] Przemysław Skurowski and Kamila Wicher. High dynamic range in x-ray imaging. In *International Conference on Information Technologies in Biomedicine*, pages 39–51. Springer, 2018. 4321

[46] Jean-Luc Starck, DL Donoho, and Michael Elad. Redundant multiscale transforms and their application for morphological component separation. *Adv. Imaging Electron Phys.*, (DAPNIA-2004-88), 2004. 4322

[47] J-L Starck, Michael Elad, and David L Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE transactions on image processing*, 14(10):1570–1582, 2005. 4322, 4323

[48] Bjarte Mehus Sunde. Early-stopping-pytorch. https://github.com/Bjarten/early-stopping-pytorch.git, 2018. 4326

[49] Donald F Swinehart. The beer-lambert law. *Journal of chemical education*, 39(7):333, 1962. 4323

[50] Jack Tumblin and Holly Rushmeier. Tone reproduction for realistic images. *IEEE Computer graphics and Applications*, 13(6):42–48, 1993. 4322

[51] David Völgyes, Anne Catrine Trægde Martinsen, Arne Stray-Pedersen, Dag Waaler, and Marius Pedersen. A weighted histogram-based tone mapping algorithm for ct images. *Algorithms*, 11(8):111, 2018. 4322

[52] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE signal processing letters*, 9(3):81–84, 2002. 4323

[53] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003. 4325

[54] Greg Ward. A contrast-based scalefactor for luminance display. *Graphics gems IV*, pages 415–421, 1994. 4322

[55] Huikai Wu, Shuai Zheng, Junge Zhang, and Kaiqi Huang. Fast end-to-end trainable guided filter. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1838–1847, 2018. 4322

[56] Zhu Xizhi. The application of wavelet transform in digital image processing. In *2008 International Conference on MultiMedia and Information Technology*, pages 326–329. IEEE, 2008. 4322

[57] R. Xu, Y. Xu, and Y. Quan. Structure-texture image decomposition using discriminative patch recurrence. *IEEE Transactions on Image Processing*, 30:1542–1555, 2021. 4323

[58] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018. 4323

[59] Chia-Hung Yeh, Chih-Hsiang Huang, and Li-Wei Kang. Multi-scale deep residual learning-based single image haze removal via image decomposition. *IEEE Transactions on Image Processing*, 29:3153–3167, 2019. 4323

[60] Xian-Shi Zhang, Kai-fu Yang, Jun Zhou, and Yong-Jie Li. Retina inspired tone mapping method for high dynamic range images. *Optics express*, 28(5):5953–5964, 2020. 4322

[61] Bolun Zheng, Shanxin Yuan, Gregory Slabaugh, and Ales Leonardis. Image demoireing with learnable bandpass filters.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3636–3645, 2020.
4322