Enhancing Dental CBCT Image by Semi-coupled Generative Adversarial Networks

Xiaotao Hu[1,2*], Zhuojia Zheng[3,8*], Huixiang Fu[4*], Yang Nan[1*], Rui Zhou[3], Kaiyu Fu[5], Dinggang Shen[6,7], Jun Xu[1]

Abstract

With the development of medical imaging technology, the Cone-Beam Computed Tomography (CBCT) has been widely used in medical diagnosis and treatment planning. Compared with Computed Tomography (CT), CBCT imaging uses lower dose but more likely produces low-resolution (LR) images with degraded visual quality. Besides, the LR CBCT images also impacts the accuracy of tooth segmentation for 3D tooth reconstruction. Since the acquired low-resolution and high-resolution (HR) images are mostly unpaired, it is difficult to directly employ supervised deep neural networks for image super-resolution. In this work, we develop an unsupervised Semi-Coupled Generative Adversarial Networks (SCGAN) to implement super-resolution (SR) of real-world CBCT images. Our SCGAN establishes two separate degradation branches to tackle the forward and backward cycle-consistent reconstruction, respectively, while being coupled by the restoration branch. This semi-coupled architecture alleviates the domain gap between unpaired LR and HR dental CBCT images. We collected a total of 6,298 dental CBCT images (slices) as the training set, including 4,038 HR dental CBCT images (slices) and 2,260 LR dental CBCT images (slices). We randomly select 500 sub-images (including crowns and roots) cropped from the LR dental CBCT images as the test set. Experiments on our collected dataset demonstrated that, our SCGAN achieves promising quantitative and qualitative SR performance, *e.g.*, Frechet Inception Distance (**FID**): 127.48; Kernel Inception Distance (**KID**): 13.99±0.28; Natural Image Quality Evaluator (**NIQE**): 4.9738, and boosts the accuracy of tooth segmentation.

Keywords

Dental CBCT image enhancement, super-resolution, semi-coupled learning, generative adversarial networks, tooth segmentation

Introduction

The Cone-Beam Computed Tomography (CBCT) imaging system developed by Arai et al. (Arai et al. 1999) and Mozzo et al. (Mozzo et al. 1998) plays an important role in oral and maxillofacial anatomy using 3-dimensional (3D) images (Kiljunen et al. 2015; Marmulla et al. 2005). CBCT system has been widely used to facilitate medical diagnosis and treatment planning, owing to its higher scanning rate and lower radiation dose over conventional computed tomography (CT) (Schulze and Drage 2020). Since CBCT imaging systems use a conical X-ray beam and a flat panel detector to perform imaging through low-dose scanning, the images are often acquired in low resolution without essential details (Bootsma et al. 2013). For example, the shape of the periodontal ligament, the morphology of the root canal and the course of hidden cracks in dental CBCT images, important for accurate dental clinics, are usually not clear enough in low-resolution images (Magalhães et al. 2022; Gao et al. 2021; Houno et al. 2017). Besides, the low-resolution of CBCT images often degrades the accuracy of tooth segmentation, which is a prerequisite step for 3D tooth reconstruction and visualization. It is therefore essential to improve the quality of dental CBCT images for better computer-aided diagnosis and treatment planning.

To improve the quality of CBCT images, one way is to capture CBCT images with higher spatial resolution, but the shooting costs and radiation dose will be increased accordingly. To keep low radiation dose while improving the quality of CBCT images, an alternative way is to employ artificial intelligence techniques. To this end, super-resolution (SR) algorithms (Dong et al. 2014; Xu et al. 2021) are widely studied to recover the corresponding high-resolution (HR) CBCT images from the input low-resolution (LR) CBCT ones (Marmulla et al. 2005).

[1]School of Statistics and Data Science, Nankai university, Tianjin, China
[2]College of Computer Science, Nankai university, Tianjin, China
[3]School of Stomatology, Tianjin Medical University, Tianjin, China
[4]School of Mathematics and Science, Nankai university, Tianjin, China
[5]Department of Pediatric Dentistry, Tianjin Stomatological Hospital, Tianjin, China
[6]School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
[7]Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200230, China
[8]Department of Orthodontics, Foshan Stomatological Hospital, Foshan, China
* Authors contributing equally to this article.
A supplemental appendix to this article is available online.

Corresponding author:
Jun Xu, School of Statistics and Data Science, Nankai University.
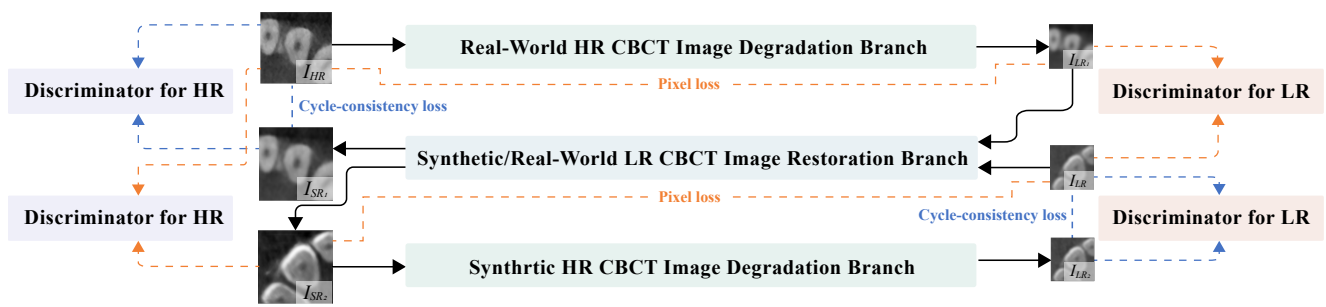Email: csjunxu@nankai.edu.cn

Figure 1. Architecture of our Semi-Coupled Generative Adversarial Network (SCGAN) for unsupervised dental CBCT image super-resolution.

With the advances of artificial intelligence on medical image analysis (Ronneberger et al. 2015; Wu et al. 2021), deep convolutional neural networks (CNNs) (LeCun et al. 2015), a branch of artificial intelligence models, have been widely exploited to advance the development of digital stomatology (Kida et al. 2018; Hwang et al. 2020; Fukuda et al. 2021). These deep learning based methods are mainly trained with amounts of paired HR and LR images to translate an LR image to its HR counterpart. To obtain paired LR and HR images, a common way is to take the given images as HR images, while the corresponding LR images are obtained by bicubic downsampling. However, the huge domain gap between the bicubic downsampling and real-world degradation often makes these methods fail on practical scenarios. Since it is difficult to obtain the corresponding HR CBCT images of the real-world LR ones, these supervised CNNs often suffer from a clear performance drop on real-world LR CBCT images.

To deal with unknown real-world degradation, unsupervised Generative Adversarial Networks (GANs) (Goodfellow et al. 2014) have been widely utilized to improve the quality of CBCT images (You et al. 2019), remove metal artifacts (Koike et al. 2020), and synthesize high-quality CT images across multiple modalities (Sun et al. 2021). The GAN often contains two sub-networks: a generator sub-network to synthesize high-quality images from low-quality ones and random noise, while the discriminator sub-network predicts whether the synthesized image is real or fake with an adversarial loss function. In recent years, many researchers have used GANs and their variants to improve CBCT image quality, and proved that these methods were capable of performing accurate SR tasks (Ledig et al. 2017; Kida et al. 2020; Bulat et al. 2018). However, due to the inevitable domain gap between the unpaired LR and HR CBCT images, the fully-cycled generative methods (Kida et al. 2020; Bulat et al. 2018) could hardly well recover the details of real-world LR CBCT images.

In this study, we aim to develop an unsupervised learning method based on Semi-Coupled Generative Adversarial Networks (SCGAN) (Hou et al. 2022) to improve the CBCT images quality, which is beneficial to show clear dental and periodontal tissues for dentists, thus achieving more accurate diagnosis. Different from CycleGAN, our SCGAN is only coupled at the LR restoration branch, while learning the cycle consistency of LR and HR CBCT image reconstructions by different degradation branches, which could better alleviate the huge domain gap between unpaired

LR and HR CBCT images. Therefore, our method can be used to exploit the unpaired LR and HR dental CBCT images for dental clinic. In addition, we compare the semantic segmentation performance of the famous UNet (Ronneberger et al. 2015) on the CBCT images before and after super-resolution. In summary, our SCGAN benefits the recovery of structures and details in LR CBCT images and improves the accuracy of the tooth segmentation.

Materials and Methods

Data Acquisition and Preprocessing

The dental CBCT images are taken from the patients underwent diagnosis and treatment at Tianjin Stomatological Hospital. Twenty cases of mid-field CBCT (80kV, 3mA, FOV: 8cm×8cm) obtained with the Morita 3D Accuitomo 170 (J Morita, Kyoto, Japan) are selected as the high-resolution (HR) samples with a voxel resolution of 0.125mm×0.125mm×0.125mm. None of the twenty patients has implants, amalgam fillings or metal crowns. Thus, the HR samples are free of metal artifacts. Another thirty cases of large-field CBCT (120kV, 10mA, FOV: 22cm×17cm) acquired by using a KaVo 3D eXam scanner (KaVo Dental, Bismarckring, Germany) are selected as the low-resolution (LR) samples with a voxel resolution of 0.3mm×0.3mm×0.3mm. The study protocol is approved by the ethics committee of the Tianjin Stomatological Hospital.

The CBCT images are then imported into the Mimics Medical 20.0 software[*] and the corresponding cross-sectional image sequences are exported for model training/testing. The images without teeth areas are removed. To this end, we obtain a total of 6,298 cross-sectional slices/images as the training set, including 4,038 HR slices/images and 2,260 LR slices/images. Each tooth in the cross-sectional images is cropped out using the labelImg labeling platform[†]. Here, we obtain the HR CBCT sub-images of size 96×96 and the LR CBCT sub-images of size 48×48. By taking the alveolar ridge of the tooth as the boundary, these sub-images are divided into the dental crown group and the dental root group. Finally, we get 49,865 HR CBCT sub-images containing 16,700 crowns and 33,165 roots, and 26,769 LR CBCT sub-images containing 9,236 crowns and 17,533 roots. We randomly select 500 samples from the LR CBCT sub-images as the test set.
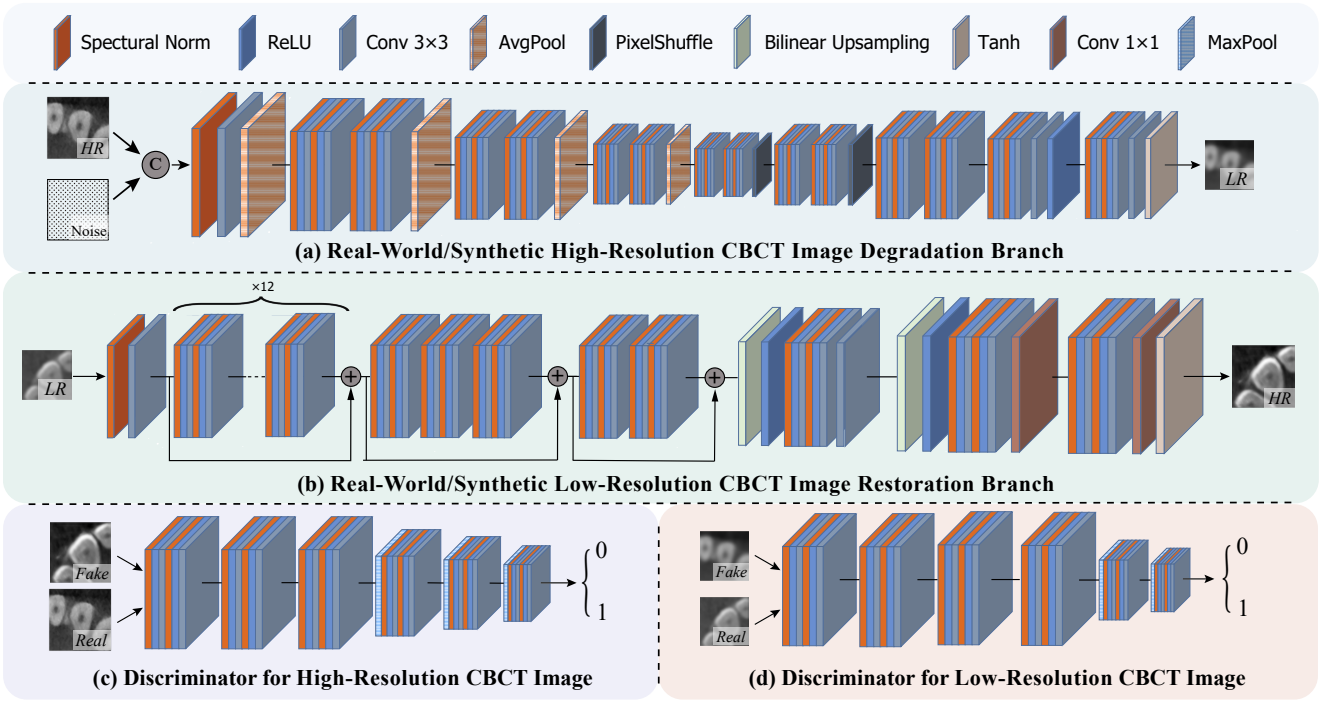

Figure 2. Architectures of the synthetic and real-world high-resolution dental CBCT image degradation branches $\mathcal{D}_1$ and $\mathcal{D}_2$ (a), the low-resolution dental CBCT image restoration branch $\mathcal{R}$ (b), and the two discriminators in (c) and (d), respectively.

Network Architecture

Here, we develop a Semi-Coupled Generative Adversarial Networks (SCGAN) to perform super-resolution of LR dental CBCT images. The overall architecture of our SCGAN is illustrated in **Figure 1**. Our SCGAN is consisted of three branches, i.e., 1) the real-world HR degradation branch $\mathcal{D}_1$, 2) the LR restoration branch $\mathcal{R}$, and 3) the synthetic HR degradation branch $\mathcal{D}_2$. Their detailed structures are visualized in **Figure 2**, together with separate structures of two discriminators used in our SCGAN.

Training Details

For network training, we take the cropped sub-images of size 48×48 and 96×96 in preprocessing procedure as the paired LR and HR dental images, respectively. The mini-batch size is set as 64. Our SCGAN is optimized by the ADAM optimizer (Kingma and Ba 2015) with β_1=0.9 and β_2=0.999 and trained for 100 epochs. The initial learning rate is set as 10^{-4} and decayed to 10^{-5} by the cosine annealing scheme (Loshchilov and Hutter 2017) for every 10 epochs. Our SCGAN is implemented in PyTorch (Paszke et al. 2019) and trained on a Tesla V100 GPU.

Performance Evaluation

Here, we compared our SCGAN with Bicubic Interpolation (BI) (Dong et al. 2014) and LRGAN (Bulat et al. 2018) on our collected dental CBCT dataset. BI is a non-parametric signal interpolation method, while LRGAN is a deep learning-based method also trained with our collected unpaired LR and HR CBCT images. To evaluate the comparison methods objectively, we use Frechet Inception Distance (FID) (Heusel et al. 2017), Kernel Inception Distance (KID) (Bińkowski et al. 2018), Natural Image Quality Evaluator (NIQE) (Mittal et al. 2012). Lower values

of these metrics indicate better performance. The calculation equations of these metrics are summarized in Appendix.

Results

The quantitative results listed in **Table 1** show that our SCGAN reaches clearly lower values (indicating better performance) in terms of FID, KID and NIQE than Bicubic Interpolation and LRGAN. This demonstrates that our SCGAN achieves better super-resolution results on the test LR CBCT images. To qualitatively evaluate Bicubic Interpolation, LRGAN, and our SCGAN, in **Figure 3**, we compare the SR results of these methods on the dental CBCT images, including crown and root groups. Taking the images in anterior area for example, our SCGAN enhances the tooth edge and root canal boundary for the images in the crown group better than Bicubic Interpolation and LRGAN. As for the root group, our SCGAN enhances the periodontal ligament of teeth and well recovers the shape of root canal. Both the quantitative and qualitative results demonstrate that our SCGAN is highly robust to real-world dental CBCT image super-resolution, and well recovers the hard-tissue structure consistent to that of real-world HR dental CBCT images. In summary, our SCGAN gains high visual quality and promising perception results on recovering the structure and details of real-world dental CBCT images.

The advantages of our SCGAN on super-resolving real-world CBCT images also brings a potential boost on semantic segmentation of dental CBCT images. To illustrate this point, we perform semantic segmentation by employing the popular UNet (Ronneberger et al. 2015) on the HR images super-resolved by Bicubic Interpolation (bicubic), LRGAN, and our SCGAN, respectively, from the original LR images. The segmentation results of the CBCT image are visualized in **Figure 4**. We observe that
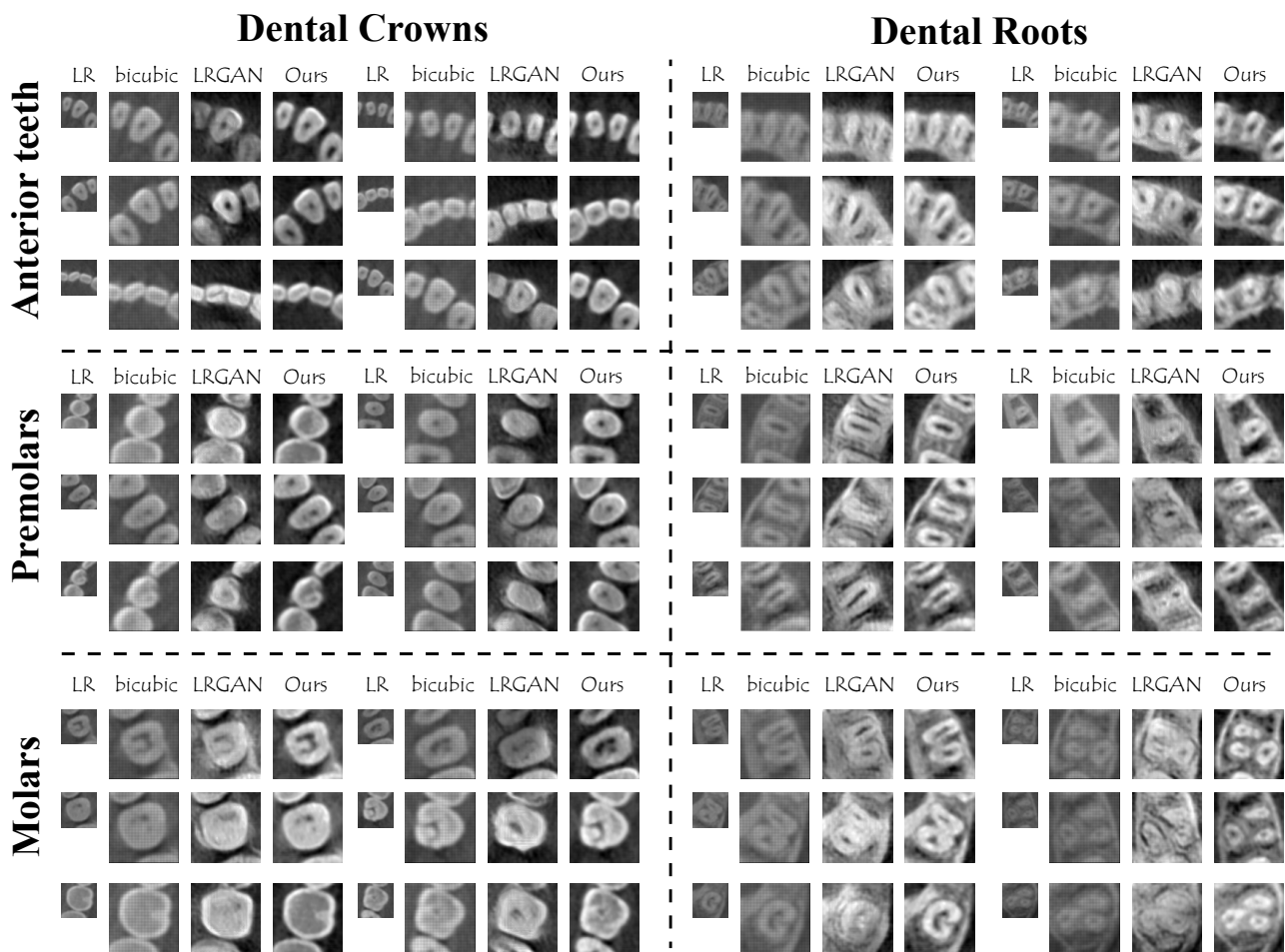
Dental Crowns
Dental Roots

Figure 3. Comparison results of super-resolution by Bicubic Interpolation, LRGAN and our SCGAN on 12 LR dental CBCT sub-images, including molar, anterior molar, and anterior tooth areas from both dental crowns and roots.

Table 1. Comparison results of FID (Heusel et al. 2017), KID (Bińkowski et al. 2018), and NIQE (Mittal et al. 2012) by Bicubic Interpolation, LRGAN (Bulat et al. 2018), and our SCGAN on dental CBCT image super-resolution. The best results are highlighted in **bold**.

Method	FID ↓	KID↓	NIQE ↓
Bicubic Interpolation	137.08	16.21±0.23	6.5827
LRGAN	134.55	15.92±0.25	5.4084
SCGAN (ours)	**127.48**	**13.99±0.28**	**4.9738**

the segmentation results of Bicubic Interpolation suffer from unclear boundaries between adjacent teeth. Besides, the segmentation results obtained by LRGAN often present inaccurate tooth boundaries. By performing super-resolution by our SCGAN, the HR CBCT images can be segmented with clear and accurate boundaries. This demonstrates that our SCGAN indeed boosts the semantic segmentation performance of U-Net on the original LR images.

Discussion

With the development of medical imaging technology, CBCT scanning has been widely used in medical diagnosis, such as oral implantology, dental and maxillofacial surgery, image-guided surgery, endodontics, periodontology and

orthodontics (Angelopoulos et al. 2012). The development of CBCT is undoubtedly a major breakthrough from 2D imaging to 3D imaging. CBCT is obviously superior to panoramic and apical radiographs in observing the positions of embedded teeth, number and shape of root canals, complex intrabony defects, and condyle-fossa relationship of temporomandibular joint, etc., which avoids the enlargement, distortion and superposition of the structure on 2D images. The 3D information embedded in CBCT images enables orthodontists to accurately evaluate complex tooth and bone malocclusion, which significantly improves dental diagnosis and treatment planning; this can not be done by traditional 2D radiography.

Early CBCT devices were specifically used for implants and dental imaging. Currently, the applied areas have been extended to the entire face and skull base. Depending on the FOV, some of the nasal cavity, paranasal sinuses, airway, cervical vertebrae and temporal bone may be displayed. In fact, specific otorhinolaryngological imaging procedures are increasingly included into CBCT systems. This suggests that CBCT may completely replaces medical CT imaging in some otolaryngology-related applications (Schulze and Drage 2020). However, compared with CT, CBCT imaging can not accurately present detailed structures, due to low irradiation dose and lack of collimator. CBCT is more likely
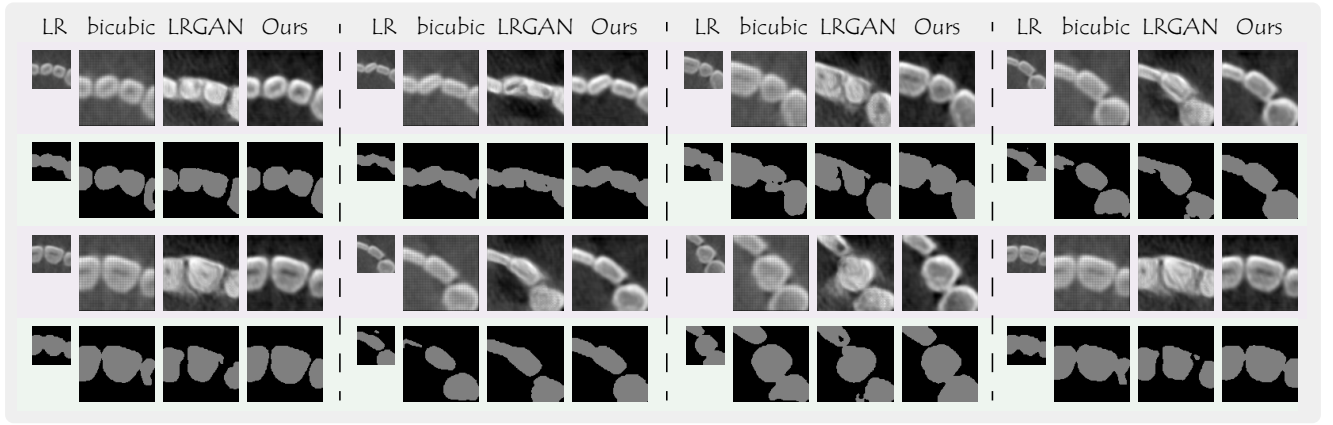
Figure 4. Comparison of semantic segmentation by UNet (Ronneberger et al. 2015) on high-resolution dental CBCT sub-images super-resolved by Bicubic Interpolation, LRGAN (Bulat et al. 2018) and our SCGAN on the low-resolution sub-images.

to produce scattering artifacts and noise that may affect the visual quality of images (Schulze and Drage 2020).

In clinical diagnosis, large-field of CBCT images in this study is suitable for the overall observation and measurement of maxillofacial soft and hard tissues by dentists. But it can not take into account the low noise of the image and the high precision in detailed area. For example, orthodontists sometimes need to judge the possibility of tooth movements after giving force according to the state of periodontal ligament, but the periodontal ligament image in large-field CBCT is often not clear and complete. Moreover, the complexity of the root canal system often brings difficulties to the root canal treatment, and the vague shape of the root canal in the large-field CBCT is not conducive to the judgment of the dentist. Therefore, it is very essential to improve the resolution and quality of dental CBCT images.

In this work, we developed a Semi-Coupled Generative Adversarial Networks (SCGAN) and evaluate it on CBCT image super-resolution. Some supervised learning methods such as deep CNNs has been considered challenging since they usually require paired LR and HR CBCT images. Generally, LR image is downsampled from the HR image which usually shows huge difference with the practical one. Our SCGAN is a more flexible solution based on unsupervised and bidirectional cycle-consistent framework Zhu et al. (2017). CycleGAN contains two fully-cycled generators for LR and HR CBCT image reconstruction. However, such fully-cycled architecture using the same branch to learn synthetic and real-world degradation process would suffer from the domain gap between the synthetic LR CBCT images and real-world ones. To address this problem, in our SCGAN we employ two separate degradation branches to learn the degradation mappings of real-world HR CBCT image and synthetic one, respectively. The two individual degradation branches in our SCGAN are only coupled at the LR image restoration branch, which also alleviates the inevitably domain gap between the LR and HR CBCT images.

In the training process of our SCGAN, the adversarial loss (Ledig et al. 2017), pixel loss (Kim et al. 2016), and cycle-consistency loss (Zhu et al. 2017) are taken into account. Specifically, the adversarial loss is used to close

the domain gap of generated SR CBCT images and real-world HR ones, and pixel loss is used to penalize the differences between the real-world HR CBCT images and the SR ones, while the cycle-consistency loss mainly contributes to preserve the structure and details. We also compare the CBCT SR performance and segmentation results of our SCGAN to that of Bicubic Interpolation and LRGAN. Bicubic Interpolation is a traditional and non-parametric method, which uses the gray values of the surrounding 16 pixels to calculate the central pixel to be found through a function. LRGAN is a deep learning-based method that only contains the forward cycle-consistent learning process for HR CBCT image restoration.

The common used metrics of FID (Heusel et al. 2017), KID (Bińkowski et al. 2018) and NIQE (Mittal et al. 2012) are employed here to validate the effectiveness of our SCGAN on dental CBCT image super-resolution objectively. The numerical results listed in **Table 1** show that, compared with Bicubic Interpolation and LRGAN, our SCGAN reaches consistently lower scores of the FID, KID and NIQE metrics, which demonstrates that the enhanced CBCT images generated by our SCGAN are with closer quality to the real-world HR dental CBCT images. The qualitative results shown in **Figure 3** demonstrate that our SCGAN can produce high-quality HR CBCT images with less scatters caused by X-rays and visible contrast in the hard-tissue structures, especially on the tooth canal boundary and alveolodental ligament.

Our SCGAN well preserves both global structure and local details in the CBCT image spatial domain. Furthermore, our SCGAN reduces the scattering artifacts and local artifacts in the reconstructed HR CBCT images. Generally, our SCGAN improves the CBCT image quality and achieves robust accurate CBCT SR performance. The quality of dental CBCT images largely influences the accuracy of tooth segmentation. We also employed the famous U-Net (Ronneberger et al. 2015) for the segmentation of dental CBCT images. The segmentation results shown in **Figure 4** illustrate that, with the help of our SCGAN, it is sufficient to recover the visible contrasts for hard-tissue diagnosis, especially the alveolar bone. The segmentation results on the SR images enhanced by Bicubic Interpolation and LRGAN show blurred boundaries between adjacent teeth, or between

periodontal tissue and jawbone. This demonstrates that these methods fail to recover the structure and integrity of the tooth contours. In conclusion, our SCGAN is superior to the baseline Bicubic Interpolation and LRGAN on recovering the overall visual quality of LR dental CBCT images, especially on the tooth structure.

To sum up, our SCGAN largely improves the visual quality of dental CBCT images and further boost the performance of dental CBCT image segmentation. Our SCGAN potentially enables the CBCT applications of adaptive treatment planning in practical dental clinics. We believe that our SCGAN can be widely used to boost the CBCT image quality for segmentation and lesion detection in clinical diagnosis. But our work also has its limitation. For example, the dataset selected in this study did not include the history of dental implants, metal crowns, other metal fillings, or maxillofacial surgery in the oral cavity. In practical clinical practice, tooth status is complicate and may suffer from artifacts. Therefore, one future research direction is to further optimize our model and make it suitable for intricate dental clinical diagnosis.

Conclusion

In this study, we have developed a Semi-Coupled Generative Adversarial Network (SCGAN) to alleviate the domain gap between unpaired LR and HR dental CBCT images. This unsupervised method contains two separate degradation branches coupled by a restoration branch, to learn forward and backward cycle-consistent reconstruction of real-world LR and HR CBCT images. Experimental results demonstrate that our SCGAN outperforms popular image super-resolution methods in recovering the structure and details of dental CBCT images, in terms of standard objective metrics and visual quality. Furthermore, with the resolution enhanced by our SCGAN, the CBCT images can be more accurately segmented for subsequent 3D tooth reconstruction. Our SCGAN can potentially reduce the radiation dose/costs spent in dental clinics and facilitates orthodontic treatment.

Author Contributions

Xiaotao Hu, contributed to experimental design, coding, experiments, and drafted the manuscript; Zhuojia Zheng, contributed to data acquisition, data annotation, and drawing figures, drafted the manuscript; Huixiang Fu, contributed to drafted the manuscript; Yang Nan, contributed to experiments and drawing figures; Rui Zhou, contributed to data acquisition and data annotation; Kaiyu Fu, contributed to data acquisition; Jun Xu, contributed to conception and experimental design, and critically revised the manuscript; Dinggang Shen, contributed to method discussion, and critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Angelopoulos C, Scarfe WC and Farman AG (2012) A comparison of maxillofacial cbct and medical ct. *Atlas of the Oral and Maxillofacial Surgery Clinics of North America* 20(1): 1–17.

Arai Y, Tammisalo E, Iwai K, Hashimoto K and Shinoda K (1999) Development of a compact computed tomographic apparatus for dental use. *Dentomaxillofacial Radiology* 28(4): 245–248.

Bińkowski M, Sutherland DJ, Arbel M and Gretton A (2018) Demystifying mmd gans. In: *International Conference on Learning Representations*.

Bootsma G, Verhaegen F and Jaffray D (2013) Spatial frequency spectrum of the x-ray scatter distribution in cbct projections. *Medical Physics* 40(11): 111901.

Bulat A, Yang J and Tzimiropoulos G (2018) To learn image super-resolution, use a gan to learn how to do image degradation first. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 185–200.

Dong C, Loy CC, He K and Tang X (2014) Learning a deep convolutional network for image super-resolution. In: *European Conference on Computer Vision*. Springer, pp. 184–199.

Fukuda M, Ariji Y, Nitoh M, Nozawa M, Kuwada C, Nishiyama M, Funakoshi T, Fujita H, Katsumata A and Ariji E (2021) Application of super-resolution convolutional neural network technique to improve the quality of soft-tissue window cone-beam ct images .

Gao A, Cao D and Lin Z (2021) Diagnosis of cracked teeth using cone-beam computed tomography: Literature review and clinical experience. *Dentomaxillofacial Radiology* 49(5): 20200407.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y (2014) Generative adversarial nets. *Advances in Neural Information Processing Systems* 27.

Heusel M, Ramsauer H, Unterthiner T, Nessler B and Hochreiter S (2017) Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems* 30.

Hou H, Hu X, Xu J, Hou Y, Wei B and Shen D (2022) Semi-cycled generative adversarial networks for real-world face super-resolution. *arXiv preprint arXiv:2205.03777* .

Houno Y, Hishikawa T, Gotoh Ki, Naitoh M, Mitani A, Noguchi T, Ariji E and Kodera Y (2017) Optimizing the reconstruction filter in cone-beam ct to improve periodontal ligament space visualization: An in vitro study. *Imaging Science in Dentistry* 47(3): 199–207.

Hwang JJ, Jung YH, Cho BH and Heo MS (2020) Very deep super-resolution for efficient cone-beam computed tomographic image restoration. *Imaging Science in Dentistry* 50(4): 331.

Kida S, Kaji S, Nawa K, Imae T, Nakamoto T, Ozaki S, Ohta T, Nozawa Y and Nakagawa K (2020) Visual enhancement of cone-beam ct by use of cyclegan. *Medical physics* 47(3): 998–1010.

Kida S, Nakamoto T, Nakano M, Nawa K, Haga A, Kotoku J, Yamashita H and Nakagawa K (2018) Cone beam computed tomography image quality improvement using a deep convolutional neural network. *Cureus* 10(4).

Kiljunen T, Kaasalainen T, Suomalainen A and Kortesniemi M (2015) Dental cone beam ct: A review. *Physica Medica* 31(8): 844–860.

Kim J, Lee JK and Lee KM (2016) Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1646–1654.

Kingma PD and Ba J (2015) Adam: A method for stochastic optimization. *International Conference on Learning Representations* .

Koike Y, Anetai Y, Takegawa H, Ohira S, Nakamura S and Tanigawa N (2020) Deep learning-based metal artifact reduction using cycle-consistent adversarial network for intensity-modulated head and neck radiation therapy treatment planning. *Physica Medica* 78: 8–14.

LeCun Y, Bengio Y and Hinton G (2015) Deep learning. *Nature* 521(7553): 436–444.

Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z et al. (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4681–4690.

Loshchilov I and Hutter F (2017) Sgdr: Stochastic gradient descent with warm restarts. *International Conference on Learning Representations* .

Magalhães J, Velozo C, Albuquerque D, Soares C, Oliveira H, Pontual ML, Ramos-Perez F and Pontual A (2022) Morphological study of root canals of maxillary molars by cone-beam computed tomography. *The Scientific World Journal* 2022.

Marmulla R, Wortche R, Muhling J and Hassfeld S (2005) Geometric accuracy of the newtom 9000 cone beam ct. *Dentomaxillofacial Radiology* 34(1): 28–31.

Mittal A, Soundararajan R and Bovik AC (2012) Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters* 20(3): 209–212.

Mozzo P, Procacci C, Tacconi A, Tinazzi Martini P and Bergamo Andreis I (1998) A new volumetric ct machine for dental imaging based on the cone-beam technique: preliminary results. *European Radiology* 8(9): 1558–1564.

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L et al. (2019) Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* 32.

Ronneberger O, Fischer P and Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, pp. 234–241.

Schulze R and Drage N (2020) Cone-beam computed tomography and its applications in dental and maxillofacial radiology. *Clinical Radiology* 75(9): 647–657.

Sun H, Lu Z, Fan R, Xiong W, Xie K, Ni X and Yang J (2021) Research on obtaining pseudo ct images based on stacked generative adversarial network. *Quantitative Imaging in Medicine and Surgery* 11(5): 1983.

Wu YH, Gao SH, Mei J, Xu J, Fan DP, Zhang RG and Cheng MM (2021) Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. *IEEE Transactions on Image Processing* 30: 3113–3126.

Xu G, Xu J, Li Z, Wang L, Sun X and Cheng MM (2021) Temporal modulation network for controllable space-time video super-resolution. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 6388–6397.

You C, Li G, Zhang Y, Zhang X, Shan H, Li M, Ju S, Zhao Z, Zhang Z, Cong W et al. (2019) Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gancircle). *IEEE Transactions on Medical Imaging* 39(1): 188–203.

Zhu JY, Park T, Isola P and Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 2223–2232.

Appendix

SCGAN Architecture

In this study, SCGAN can be divided into three branches, namely 1) the real-world HR CBCT image degradation branch $\mathcal{D}_1$, 2) the LR CBCT image restoration branch $\mathcal{R}$, and 3) the synthetic HR degradation branch $\mathcal{D}_2$. We set $\alpha=1$, $\beta=0.05$ in Eqs. (1)~(9) and $\theta=1$, $\gamma=0.05$ in Eq. (4).

Synthetic Degradation Branch on HR CBCT Image

This branch, denoted as $\mathcal{D}_1$, aims to degrade a real-world HR image I_{HR} to a synthetic LR one. It is the degradation stage of the forward cycle-consistency learning process "$I_{HR} \rightarrow \mathcal{D}_1(I_{HR}) \rightarrow I_{LR_1}$" in which the corresponding restoration stage is implemented by the LR image restoration branch $\mathcal{R}$ introduced as follows. The HR image with noise is first used as the input of $\mathcal{D}_1$. Noise is used to simulate the image degradation process of different degrees in real situations in a more diverse way. The output of $\mathcal{D}_1$ is the synthetic LR image I_{LR_1} corresponding to the input I_{HR_1}. The overall loss function for this branch $\mathcal{D}_1$ is

$$l_{\mathcal{D}_1} = \alpha l_{adv}^{\mathcal{D}_{L1}} + \beta l_{pixel}^{I_{LR_1}} \tag{1}$$

where α and β are the weights of the two loss functions.

The adversarial loss $l_{adv}^{\mathcal{D}_{L1}}$ uses the discriminator $\mathcal{D}_{L_1}$ to predict the real-world LR image I_{LR} as 1 and the synthetic LR one I_{LR_1} as 0, respectively. We use the hinge loss as follows:

$$l_{adv}^{\mathcal{D}_{L_1}} = \underset{I_{rL} \sim P_{rL}}{E}[\min(0, -1 + \mathcal{D}_{L_1}(I_{LR}))] +$$
$$\underset{I_{sL} \sim P_{sL}}{E}[\min(0, -1 - \mathcal{D}_{L_1}(I_{LR_1}))] \tag{2}$$

where P_{rL} and P_{sL} are the distributions of real-world LR images I_{LR} and the synthetic one I_{LR_1} degraded by $\mathcal{D}_1$ from the real-world HR CBCT image I_{HR}, respectively. $\mathcal{D}_{L_1}$ is the discriminator.

The pixel loss $l_{pixel}^{I_{LR_1}}$ is calculated between the synthetic degradation image I_{LR_1} and the input HR image I_{HR} downsampled to the same resolution with I_{LR_1} by average pooling, which are computed as follows,

$$l_{pixel}^{I_{LR_1}} = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} |F_1(I_{HR})_{i,j} - F_2(I_{LR_1})_{i,j}| \tag{3}$$

W, H are the width and height of the generated image I_{LR_1}, F_1 and F_2 map I_{HR} and I_{LR_1} to the same resolution, respectively. Therefore, an l_1 loss can be calculated pixel by pixel, where F_1 is implemented by average pooling and F_2 is an identity mapping.

Restoration Branch on LR CBCT Image

The LR CBCT image restoration branch $\mathcal{R}$ is shared by the forward and backward cycle-consistency learning processes. In the forward learning process, it restores a synthetic LR image I_{LR_1} degraded from the HR image I_{HR} via $\mathcal{D}_1$, while in the backward learning process, it restores the real-world LR image I_{LR} to a synthetic HR image. Here, we denote the synthetic HR image restored from I_{LR_1} as I_{SR_1} and the one restored from I_{LR} as I_{SR_2}.

The restoration branch $\mathcal{R}$ aims to generate high-quality SR image, shared by the forward and backward learning processes. For the loss function, we use both adversarial loss $l_{adv}^{\mathcal{D}_1}$ and cycle-consistency loss $l_{cycle}^{I_{SR_1}}$ in the forward learning process, and use the combination of adversarial loss $l_{adv}^{\mathcal{D}_2}$ and pixel loss $l_{pixel}^{I_{SR_2}}$ in the backward learning process. The overall loss function for this branch is

$$l_{\mathcal{R}} = \theta l_{\mathcal{R}}^{I_{SR_1}} + \gamma l_{\mathcal{R}}^{I_{SR_2}} \tag{4}$$

where θ and γ are the corresponding weights, and

$$l_{\mathcal{R}}^{I_{SR_1}} = \alpha l_{adv}^{\mathcal{D}_1} + \beta l_{cycle}^{I_{SR_1}} \tag{5}$$

$$l_{\mathcal{R}}^{I_{SR_2}} = \alpha l_{adv}^{\mathcal{D}_2} + \beta l_{pixel}^{I_{SR_2}} \tag{6}$$

Degradation Branch on Real-World HR CBCT Image

This branch, denoted as $\mathcal{D}_2$, learns to degrade the SR CBCT scans I_{SR_2} restored from the real-world LR image I_{LR_2} via $\mathcal{R}$. Since the real-world and synthetic LR images suffer from an inevitable degradation gap, it is reasonable to use different branches for degrading the real-world HR image I_{HR} and the synthetic one I_{SR_1} generated from the restoration branch $\mathcal{R}$. The backward cycle-consistent learning process can be describe as "$I_{LR} \rightarrow \mathcal{R}(I_{LR}) \rightarrow \mathcal{D}_2(\mathcal{R}(I_{LR}))$". The overall loss function for this branch is

$$l_{\mathcal{D}_2} = \alpha l_{adv}^{\mathcal{D}_{L_2}} + \beta l_{cycle}^{I_{LR_2}} \tag{7}$$

The adversarial loss $l_{adv}^{\mathcal{D}_{L_2}}$ uses the discriminator $\mathcal{D}_{L_2}$ to predict the real-world LR image I_{LR} as 1 and the synthetic LR one I_{LR_1} as 0. We use the hinge loss as follows:

$$l_{adv}^{\mathcal{D}_{L_2}} = \mathop{E}_{I_{rL} \sim P_{rL}} [\min(0, -1 + \mathcal{D}_{L_2}(I_{LR}))] + \\ \mathop{E}_{I_{sL} \sim P_{sL}} [\min(0, -1 - \mathcal{D}_{L_2}(I_{LR_2}))] \tag{8}$$

where P_{rL} and P_{sL} are the distributions of real-world LR images I_{LR} and synthetic one I_{LR_2} degraded by $\mathcal{D}_2$ from the real-world SR CBCT image I_{SR_2}, respectively. $\mathcal{D}_{L_2}$ is the discriminator.

The cycle-consistency loss l_{cycle} is an l_1 loss function used here to well recover the details the via the restoration branch $\mathcal{R}$:

$$l_{cycle}^{I_{LR_2}} = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} |(I_{LR})_{i,j} - (I_{LR_2})_{i,j}| \tag{9}$$

Evaluation Metrics

On the test sets, we use the following metrics: Frechet Inception Distance (FID), Kernel Inception Distance (KID), and Natural Image Quality Evaluator (NIQE) to evaluate the visual quality of dental CBCT images.

Frechet Inception Distance (FID) is a commonly used metric to evaluate the distribution distance between the real-world and SR images. In our experiment, the Inception Net-V3 pretrained network uses the last pooling layer as the output layer, which outputs a 2048-dimensional feature vector. FID is obtained by the mean and covariance difference of the feature vectors of the two sets of images:

$$FID(x, g) = \|\mu_x - \mu_g\| + Tr(\Sigma_x + \Sigma_g - 2\sqrt{\Sigma_x \Sigma_g}) \tag{10}$$

where μ_x, μ_g, and Σ_x, Σ_g represent the mean and covariance matrices of feature vectors output by the real image set and the generated image set via Inception Net-V3, respectively.

Kernel Inception Distance (KID) is developed by Maximum Mean Discrepancy (MMD). MMD is an integral probability metric (integral probability metrics), that is, finding a continuous function f in the sample space, then calculating the means of the function values of samples with different distributions on f. By taking the difference of the two means, the mean discrepancy of the two distributions corresponding to f can be obtained. KID is equal to the square of Maximum Mean Discrepancy MMD^2 between Inception representation, namely:

$$MMD_u^2(X, Y) = \frac{1}{m(m-1)} \sum_{i \neq j}^{m} k(x_i, x_j) + \\ \frac{1}{n(n-1)} \sum_{i \neq j}^{n} k(y_i, y_j) - \frac{2}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} k(x_i, y_j) \tag{11}$$

where kernel $k(x, y) = \left(\frac{1}{d} x^\top y + 1\right)^3$. X represents the distribution of generated images, Y represents the distribution of real images m is the sample size of the generated image, and n is the sample size of the real image. The higher KID, the better the performance of the generative model. Compared to FID, KID has no bias and small variance. Besides, it needs fewer samples and computationally faster than FID.

Natural Image Quality Evaluator (NIQE). In our experiment, we model a series of quality-aware features with a multivariate Gaussian MVG model. NIQE measures the quality of a lossy image is the distance between its MVG and the MVG of a natural image, namely:

$$A = \nu_1 - \nu_2, \tag{12}$$

$$Dist(A, \Sigma_1 \Sigma_2) = \left(A^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} A\right)^{\frac{1}{2}} \tag{13}$$

where ν_1, ν_2 and Σ_1, Σ_2 represent the means and covariance matrices of the natural MVG model and the distorted image's MVG model.