Task Based Losses

Kamal Nambiar

Computational Medical Imaging
Ferienakademie

27 September 2018

Loss Functions

- Loss function is a quantitative comparison between the predicted output of a model and the expected output.
- The magnitude of the loss function tells how 'bad' the model parameters are.

$$
\begin{aligned}
\theta^* &= \arg\min_{\theta} \mathcal{L}(\theta) + \lambda \cdot \Phi(\theta) \\
&= \frac{1}{n}\sum_{i=1}^{n} \arg\min_{\theta} \mathcal{L}(y^{(i)}, f(x^{(i)};\theta)) + \lambda \cdot \Phi(\theta)
\end{aligned}
$$

Examples

- Mean Squared Error (MSE): Commonly used in regression problems

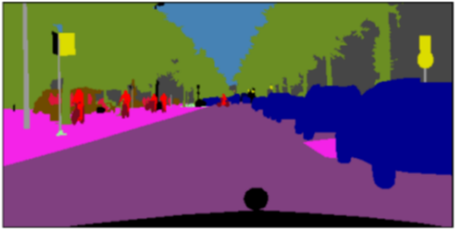
$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)})^2$$

- Cross Entropy Loss: Commonly used in classification problems

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}\sum_{c} -[y_c^{(i)} log(\hat{y_c}^{(i)})]$$

Photographic Image Synthesis with Cascaded Refinement Networks Chen and Koltun 2017

Objective: Given semantic specification of the scene, produce the corresponding photographic image



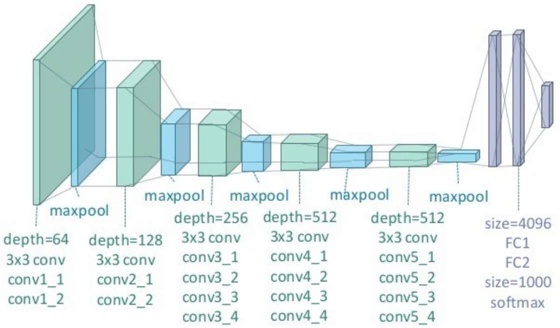
(a) Input: Semantic layout of a road scene

(b) Output: Synthesized Photographic Image

Photographic Image Synthesis with Cascaded Refinement Networks

- Inverse Semantic Segmentation problem.
- One to Many problem; no absolute ground truth.
- The output image from the training set is treated as a reference image.
- A loss function involving pixel by pixel comparison will result in large losses when the network returns one of the other possible solutions of the problem
- A loss function that compares the features between the predicted image and reference image

VGG-19 Network



- Trained on the ImageNet Dataset consisting of 14M images across 20K classes.
- 'conv1_2', 'conv2_2','conv3_2', 'conv4_2', and 'conv5_2' in VGG-19 for calculating loss

Perceptual Loss

- The idea is to match activations in a visual perception network that is applied seperately to the network output and the reference image
- Visual Perception Network refers to a pretrained classification neural network (VGG-19)
- Layers of the network represent increasing layers of abstraction. This helps to compare the features of the two images at different levels of granularity.
- For a training pair (I,L), with Φ_ℓ is the activation at layer ℓ, and g is the image synthesis network that is to be trained, the loss is given by:

$$\mathcal{L}_{I,L}(\theta) = \sum_{\ell} \lambda_\ell \|\Phi_\ell(I) - \Phi_\ell(g(L;\theta))\|_1$$

Result Comparison

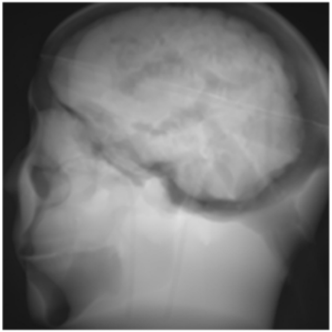
(a) Input: Semantic Layout of Room

(b) Output: Proposed Network

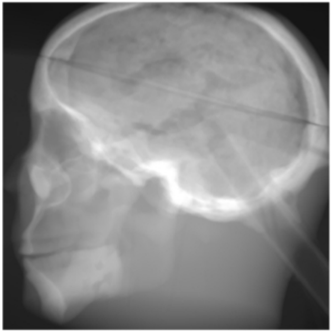
(c) Output: cGAN

MR to X-ray Projection Image Synthesis Stimpel et al. 2017

Objective: Apply Deep learning-based methods for X-ray projection image synthesis from MR projections



MR proj.

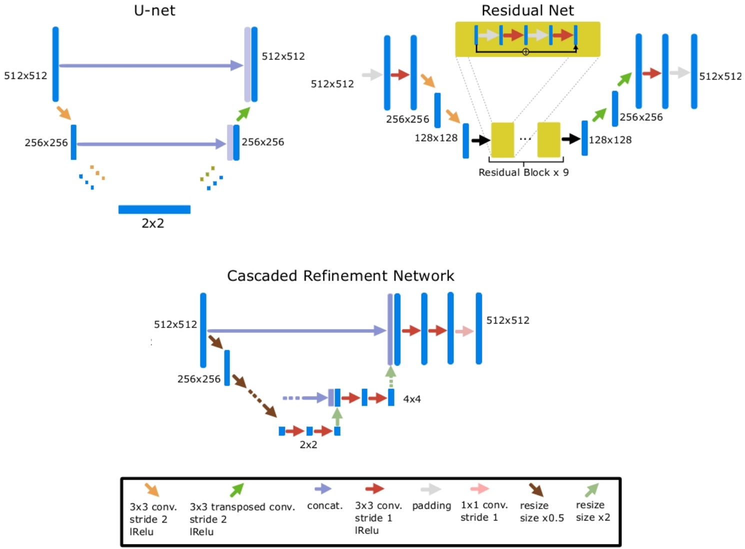


X-ray proj.

Perceptual Loss

- The network from the Photographic Image Synthesis with Cascaded Refinement Networks paper Chen and Koltun 2017 is used but instead of semantic images, MR images are used as network input. The network is trained using CT images as output.
- The performance of the network is compared against UNet and ResNet.
- The performance of the perception loss function is compared against l_1 loss.

Networks used for comparison

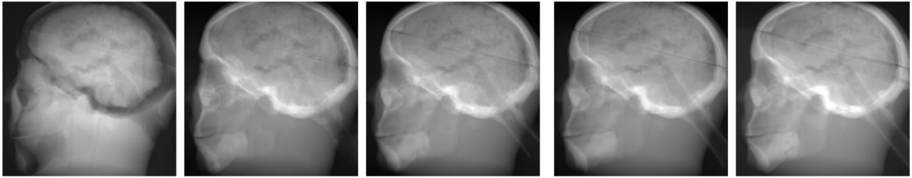


Quantitative results

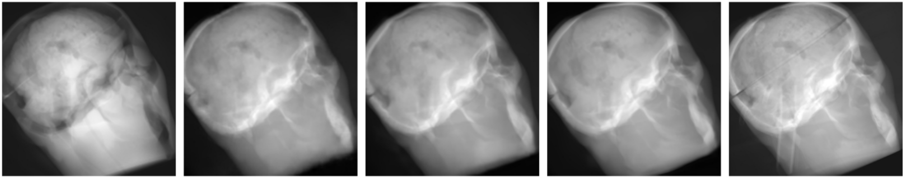
Network - loss	MAE	SSIM	PSNR
UNet - p loss	0.083	0.891	26.994
ResNet - p loss	0.077	0.924	27.675
CRN - p loss	**0.071**	**0.931**	**28.353**
UNet - l_1 loss	0.068	0.917	28.506
ResNet - l_1 loss	**0.058**	**0.938**	**30.067**
CRN - l_1 loss	0.084	0.92	27.097

Table: Comparing image metrices like Mean Absolute Error (MAE), Structural Similarity Index Measure (SSIM) and Peak Signal to Noise Ratio (PSNR) for various network architecture and loss functions

Result Comparison



(a) Input: MR proj. (b) Output: U-net p-loss. (c) Output: ResNet p-loss. (d) Output: CRN p-loss. (e) Reference: X-ray proj.

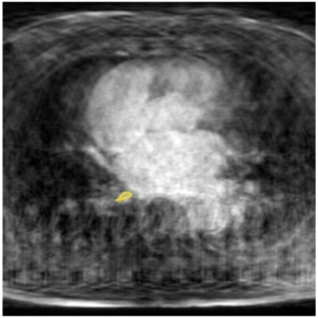


(f) Input: MR proj. (g) Output: U-net ℓ_1-loss. (h) Output: ResNet ℓ_1-loss. (i) Output: CRN ℓ_1-loss. (j) Reference: X-ray proj.

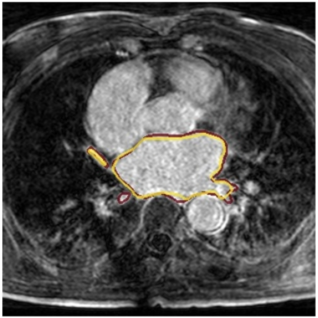
The perceptual-loss conserves even small high-frequency details in image-to-image transfer. Because high-spatial resolution is desired in most fluoroscopic procedures, using perceptual-loss function for the underlying task produces best results.

Adversarial and Perceptual Refinement for Compressed Sensing MRI Reconstruction Seitzer et al. 2018

Objective: Optimally combine Adversarial and Perceptual loss functions with the MSE loss function for compressed sensing-based MR Imaging

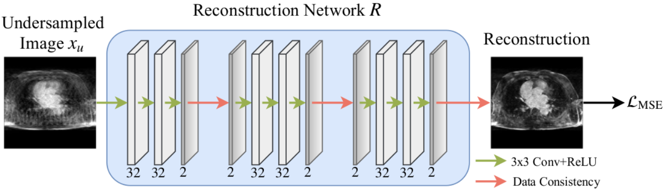


Zero-filled

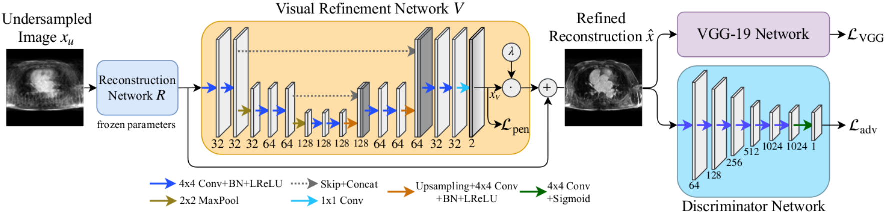


Ground Truth

Proposed Network



(a) Stage 1: training of reconstruction network using $\mathcal{L}_{\text{MSE}}(R)$.



(b) Stage 2: training of visual refinement network using $\mathcal{L}_{\text{vis}}(V)$.

Loss Function for Stage 2

- Stage 1 network is trained for MSE loss
- Stage 2 network loss is given by:

$$\mathcal{L}(V) = \frac{1}{2}\left(\frac{\mathcal{L}_{\text{adv}}(V)}{M} + \frac{\mathcal{L}_{\text{feat}}(V)}{N}\right) + \frac{\mathcal{L}_{\text{VGG}}(V)}{O} + \alpha\mathcal{L}_{\text{pen}}(V)$$

$$\mathcal{L}_{\text{pen}}(V) = \|x_V\|_1$$

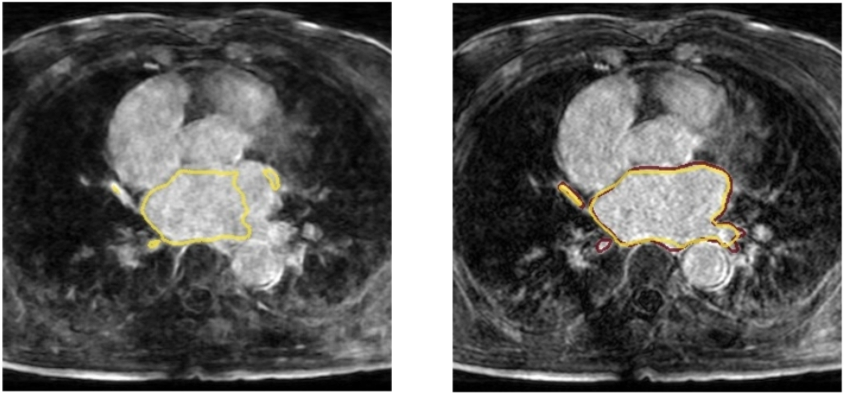
- M , N , O constants set such that $\frac{\mathcal{L}_{adv}(V)}{M} = \frac{\mathcal{L}_{feat}(V)}{N} = \frac{\mathcal{L}_{VGG}(V)}{O} = 1$ in the first iteration of training, which amounts to assigning the two adversarial loss terms the same initial importance as $\mathcal{L}_{VGG}(V)$.
- The penalty strength α is important for training speed and stability. Choosing α such that $\mathcal{L}_{\text{pen}}(V) \approx 0.1$ at first iteration

Quantitative results

Method	PSNR	MOS	SIS
Ground Truth	∞	3.78 ± 0.45	1
RecNet	$\mathbf{32.46 \pm 2.26}$	2.75 ± 0.78	0.801
DLMRI	31.45 ± 2.40	1.09 ± 0.29	0.842
DAGAN	28.41 ± 1.91	2.61 ± 0.83	0.812
Proposed Model	31.89 ± 2.18	$\mathbf{3.24 \pm 0.63}$	**0.941**

Table: Comparing image metrices like Peak Signal to Noise Ratio (PSNR), Semantic Interpretability Score (SIS) and mean opinion score (MOS) for various network architectures

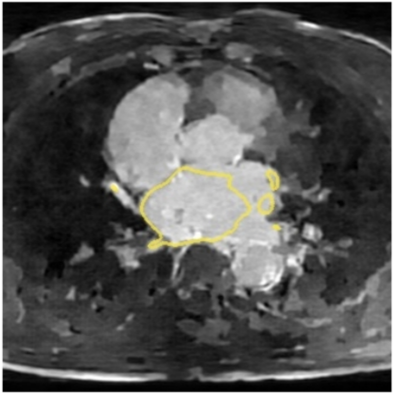
Result Comparison



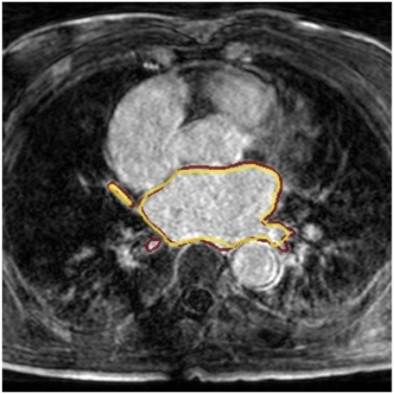
RecNet Ground Truth

RecNet Schlemper et al. 2018: Similar to the proposed approach but without the stage 2-refinement step

Result Comparison



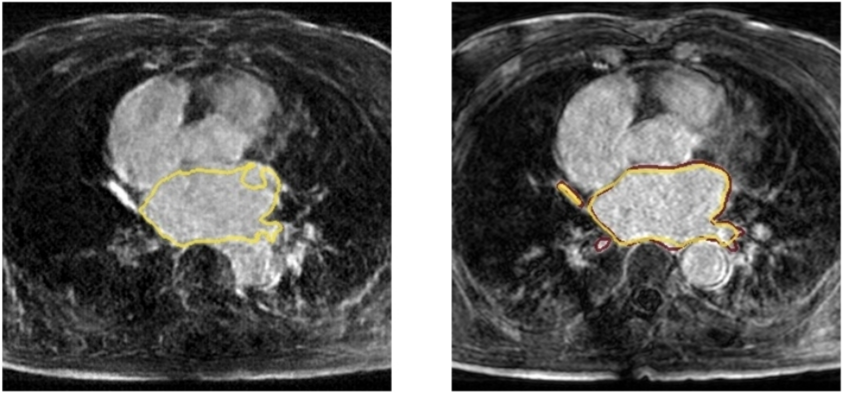
DLMRI



Ground Truth

DLMRI Ravishankar and Bresler 2011: A dictionary learning based method.

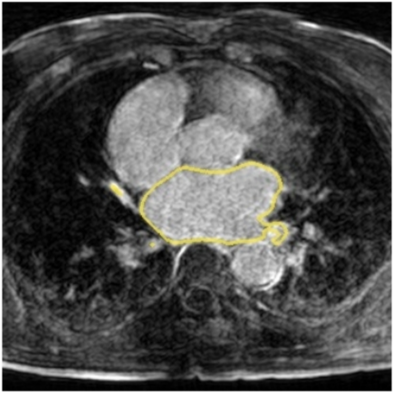
Result Comparison



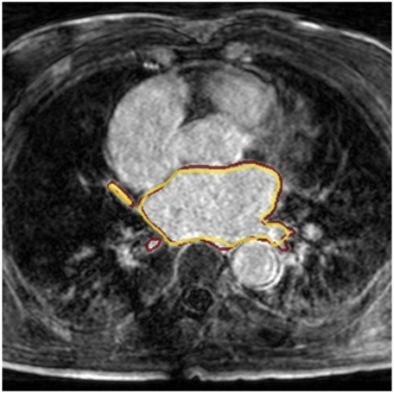
DAGAN Ground Truth

DAGAN Yang et al. 2018: Combines MSE loss with a visual loss function without any further precautions

Result Comparison



Proposed Method



Ground Truth

References I

Chen, Q. and V. Koltun (2017). “Photographic Image Synthesis with Cascaded Refinement Networks”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 1520–1529. DOI: 10.1109/ICCV.2017.168.

Ravishankar, S. and Y. Bresler (2011). “MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning”. In: *IEEE Transactions on Medical Imaging* 30.5, pp. 1028–1041. ISSN: 0278-0062. DOI: 10.1109/TMI.2010.2090538.

Schlemper, J. et al. (2018). “A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction”. In: *IEEE Transactions on Medical Imaging* 37.2, pp. 491–503. ISSN: 0278-0062. DOI: 10.1109/TMI.2017.2760978.

Seitzer, M. et al. (2018). “Adversarial and Perceptual Refinement for Compressed Sensing MRI Reconstruction”. In: *ArXiv e-prints*. arXiv: 1806.11216 [cs.CV].

References II

Stimpel, B. et al. (2017). "MR to X-Ray Projection Image Synthesis". In: *ArXiv e-prints*. arXiv: 1710.07498 [cs.CV].

Yang, G. et al. (2018). "DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction". In: *IEEE Transactions on Medical Imaging* 37.6, pp. 1310–1321. ISSN: 0278-0062. DOI: 10.1109/TMI.2017.2785879.

Thank you for listening
Any questions?