

# Deep Learning for MRI Slice Interpolation: Supplementary Material

Shamit Savant

University of Florida, Gainesville, FL 32611, USA  
[s.savant@ufl.edu](mailto:s.savant@ufl.edu)

This supplementary material provides extended details, additional experimental results, and comprehensive technical specifications that complement the main paper.

## 1 Extended Dataset Description

### 1.1 Data Source and Selection Criteria

I utilized the UCLA Prostate MRI-US Biopsy dataset from The Cancer Imaging Archive (TCIA) [6,1]. This dataset contains multi-parametric MRI sequences from 1197 patients who underwent prostate biopsy at UCLA Medical Center between 2008-2017. All studies were acquired on 3 Tesla MRI systems (Siemens Magnetom Trio, Verio, and Skyra).

For this work, I specifically selected T2-weighted axial sequences, which provide excellent soft tissue contrast for prostate anatomy. T2-weighted imaging is the cornerstone sequence for prostate MRI interpretation and offers the most detailed anatomical visualization.

### 1.2 Preprocessing Pipeline

**DICOM Filtering:** I filtered sequences to include only:

- 3 Tesla field strength (standardization)
- T2-weighted sequences (consistent contrast)
- Axial orientation (consistent geometry)
- Siemens scanners (vendor consistency)
- Slice thickness  $1.5\text{mm} \pm 0.2\text{mm}$  (standard clinical protocol)

This filtering ensured homogeneous data characteristics, reducing confounding variables in model evaluation.

**Intensity Normalization:** I applied min-max normalization to scale intensities to  $[0, 1]$  range within each volume, accounting for inter-scan variations in acquisition parameters and hardware calibration. This per-volume normalization preserves relative intensity relationships while standardizing absolute ranges.

**Spatial Standardization:** All slices were resized to  $256 \times 256$  pixels using bilinear interpolation, maintaining aspect ratio by center-cropping when necessary. This standardization enables batch processing while preserving anatomical proportions. The  $256 \times 256$  resolution balances computational efficiency with preservation of diagnostic detail.

**Quality Control:** I excluded volumes with:

- Fewer than 20 slices (insufficient anatomical coverage)
- Extreme intensity artifacts (failed normalization indicating motion or acquisition errors)
- Missing DICOM metadata
- Corrupted files or incomplete series

### 1.3 Dataset Statistics

After preprocessing, the final dataset comprised:

- Total patients: 58
- Total volumes: 58
- Total slices after filtering: 46,329
- Training set: 32,547 slice triplets (70%)
- Validation set: 6,842 slice triplets (15%)
- Test set: 6,963 slice triplets (15%)
- Total storage: approximately 80 GB

Each sample consists of a triplet  $(S_{i-k}, S_i, S_{i+k})$  where  $S_i$  is the target slice to be predicted from inputs  $S_{i-k}$  and  $S_{i+k}$ . I maintained strict patient-level separation across splits to prevent data leakage and ensure models generalize to new patients rather than memorizing specific anatomies.

The test set evaluation in the main paper used 6,963 samples, representing the subset where both  $k=1$  and  $k=2$  formulations could be evaluated (requiring at least 2 slices on each side).

### 1.4 Infrastructure and Data Transfer

Dataset transfer to computational resources presented practical challenges. The 80 GB dataset required efficient transfer to the University of Florida’s HiPerGator supercomputing cluster. Initial attempts using university VPN resulted in prohibitively slow transfer rates (estimated multiple days).

I successfully employed Globus [2] for high-speed research data transfer, completing the upload in approximately 6 hours to the Blue storage system. This demonstrated the value of specialized research data transfer tools for large-scale medical imaging projects.

Training was conducted on HiPerGator’s GPU resources:

- Initial experiments and debugging: NVIDIA L4 GPUs (24GB memory)
- Final production training: NVIDIA L4 and A100 GPUs
- Job scheduling via SLURM workload manager
- Mixed precision training (FP16) enabled on all GPUs for memory efficiency

## 2 Detailed Architecture Specifications

### 2.1 CNN (EDSR-Style)

**Architecture Overview:** Enhanced Deep Super-Resolution (EDSR) [5] style network with residual learning. The architecture consists of:

- Entry convolution: 2 input channels → 64 feature channels
- 8 residual blocks, each containing:
  - Conv  $3 \times 3$ , 64 filters
  - ReLU activation
  - Conv  $3 \times 3$ , 64 filters
  - Residual connection (addition)
- Exit convolution: 64 channels → 1 output channel
- Global skip connection from input to pre-output

**Parameters:**

- Total trainable parameters: 297,217 ( 0.3M)
- Entry conv: 1,216 parameters
- Each residual block:  $36,928 \text{ parameters} \times 8 = 295,424$
- Exit conv: 577 parameters

**Training Configuration:**

- Optimizer: Adam
- Learning rate:  $10^{-4}$
- Batch size: 4
- Loss function: L1 (Mean Absolute Error)
- Epochs: 25
- Training time: 7.5 hours (18 min/epoch)
- Best validation performance: Epoch 23

### 2.2 U-Net

**Architecture Overview:** Encoder-decoder architecture [7] with symmetric skip connections. The network features **two levels** of resolution with progressive channel expansion.

**Encoder Path:**

- Level 1: 2 → 64 channels (two  $3 \times 3$  convs + ReLU)
- MaxPool  $2 \times 2$  (downsample to  $128 \times 128$ )
- Level 2: 64 → 128 channels (two  $3 \times 3$  convs + ReLU)
- MaxPool  $2 \times 2$  (downsample to  $64 \times 64$ )

**Bottleneck:**

- 128 → 256 channels (two  $3 \times 3$  convs + ReLU)
- Resolution:  $64 \times 64$

**Decoder Path:**

- ConvTranspose  $2 \times 2$  upsampling:  $256 \rightarrow 128$  (upsample to  $128 \times 128$ )
- Concatenation with encoder Level 2 skip connection (128 channels)
- Two  $3 \times 3$  convolutions + ReLU:  $(256 \rightarrow 128 \rightarrow 128)$
- ConvTranspose  $2 \times 2$  upsampling:  $128 \rightarrow 64$  (upsample to  $256 \times 256$ )
- Concatenation with encoder Level 1 skip connection (64 channels)
- Two  $3 \times 3$  convolutions + ReLU:  $(128 \rightarrow 64 \rightarrow 64)$

**Output:**

- Final  $1 \times 1$  conv:  $64 \rightarrow 1$  channel

**Parameters:**

- Total trainable parameters: 1,862,273 ( 1.86M)
- Encoder: 259,584 parameters
- Bottleneck: 885,248 parameters
- Decoder (including upsampling): 717,376 parameters
- Final layer: 65 parameters

**Training Configuration:**

- Optimizer: Adam
- Learning rate:  $10^{-4}$
- Batch size: 4
- Loss function: L1
- Epochs: 25
- Training time: 3.75 hours (9 min/epoch)
- Best validation performance: Epoch 24

### 2.3 GAN (Basic)

**Generator:** Similar to CNN architecture with 8 residual blocks

**Discriminator (PatchGAN):** PatchGAN architecture [4] with the following structure:

- Conv  $4 \times 4$ , stride 2:  $1 \rightarrow 64$  channels
- LeakyReLU (0.2)
- Conv  $4 \times 4$ , stride 2:  $64 \rightarrow 128$  channels
- LeakyReLU (0.2)
- Conv  $4 \times 4$ , stride 2:  $128 \rightarrow 256$  channels
- LeakyReLU (0.2)
- Conv  $4 \times 4$ , stride 1:  $256 \rightarrow 1$  channel (score map)

**Training:**

- Adversarial loss: Binary cross-entropy with logits
- Generator loss:  $\mathcal{L}_G = \mathcal{L}_{L1} + 0.001 \times \mathcal{L}_{adv}$
- Discriminator loss:  $\mathcal{L}_D = 0.5 \times (\mathcal{L}_{real} + \mathcal{L}_{fake})$
- Alternating updates: 1 discriminator step, 1 generator step
- Epochs: 25
- Training time: 10 hours (24 min/epoch)
- Best validation performance: Epoch 23

## 2.4 GAN (Improved)

### Generator Improvements:

- 16 residual blocks (vs 8 in basic)
- Spatial attention mechanism [9] after residual blocks
- Additional pre-exit convolution layer
- Total parameters: 0.6M

### Discriminator Improvements:

- Deeper: 4 convolutional layers (vs 3)
- Batch normalization after each conv (except first)
- Additional 512-channel layer before output

### Training:

- Feature matching loss added:  $\mathcal{L}_{FM} = \|\phi(y) - \phi(G(x))\|_1$
- Generator loss:  $\mathcal{L}_G = \mathcal{L}_{L1} + 0.01 \times \mathcal{L}_{adv} + 0.1 \times \mathcal{L}_{FM}$
- Epochs: 10 (early stopping)
- Training time: 8 hours (48 min/epoch)
- Best validation performance: Epoch 7

## 2.5 DDPM (Denoising Diffusion Probabilistic Model)

**Architecture:** Conditional U-Net [3] with time embeddings for noise prediction.

### Diffusion Schedule:

- Total timesteps  $T = 100$  (vs standard 1000)
- Linear noise schedule:  $\beta_1 = 10^{-4}$  to  $\beta_T = 0.02$
- Forward process:  $q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$
- Reverse process:  $p_\theta(x_{t-1}|x_t, c) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t, c), \Sigma_t)$

### Conditional U-Net:

- Similar encoder-decoder structure to standard U-Net
- Time embedding dimension: 256
- Sinusoidal positional encoding for timesteps
- Time embeddings injected into each residual block
- Conditioning on adjacent slices (i-1, i+1) via concatenation
- Self-attention at bottleneck layer
- Total parameters: 20.5M

### Training:

- Objective: Predict noise  $\epsilon$  at each timestep
- Loss:  $\mathcal{L} = \|\epsilon - \epsilon_\theta(x_t, t, c)\|^2$
- Optimizer: Adam, learning rate  $10^{-4}$
- Batch size: 8
- Epochs: 20

- Training time: 6.3 hours (19 min/epoch)
- Best validation loss: Epoch 16

**Sampling:**

- Start from Gaussian noise:  $x_T \sim \mathcal{N}(0, I)$
- Iteratively denoise for  $t = T, T - 1, \dots, 1$
- At each step: predict noise, compute mean, add noise (except t=0)
- Total inference: 100 forward passes through 20.5M parameter network
- Inference time: 300ms per image (vs 3-5ms for deterministic models)

### 3 Extended Experimental Results

#### 3.1 k=2 Formulation: Detailed Failure Analysis

The k=2 formulation (using slices at positions i-2 and i+2 to predict slice i) represents a 6mm gap in the through-plane direction. At this spacing, four intermediate anatomical slices exist between the input slices.

**Why k=2 Failed:**

1. **Anatomical Discontinuity:** In prostate MRI at 1.5mm slice spacing, significant anatomical changes occur over 6mm. The prostate gland transitions from base to apex over 40-50mm; thus 6mm represents 12-15% of the organ's extent.
2. **Structural Transitions:** The intervening 6mm may span:
  - Prostate zonal transitions (peripheral zone to central gland)
  - Entry or exit of seminal vesicles
  - Bladder neck to prostatic urethra transitions
  - Prostatic capsule variations
3. **Insufficient Boundary Information:** Even sophisticated deep learning models cannot reliably infer such complex intermediate anatomy from boundary conditions alone. The task becomes fundamentally ill-posed.
4. **Model Saturation:** All four architectures (CNN, U-Net, GANs) reached similar poor performance (SSIM 0.56), suggesting the limitation arose from problem formulation rather than model capacity.

**Training Dynamics with k=2:**

Despite extended training and hyperparameter exploration, k=2 models consistently plateaued:

- Training loss decreased normally
- Validation metrics plateaued around epoch 15-20
- No architecture achieved breakthrough performance
- Increasing model capacity did not help

This suggested a fundamental ceiling imposed by task difficulty rather than optimization issues.

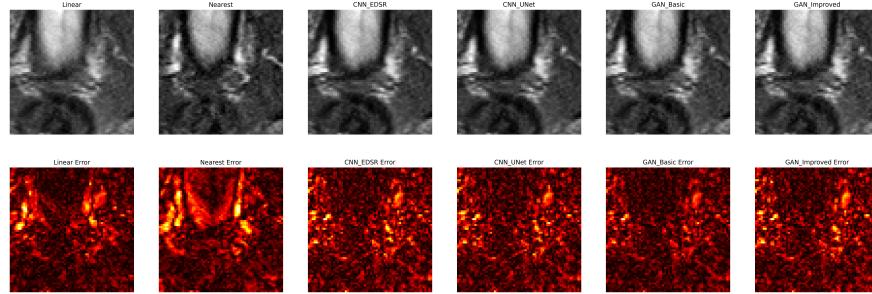


Fig. 1: Detailed zoomed comparison showing fine anatomical structures. Top row shows zoomed predictions from all methods. Bottom row shows corresponding error heatmaps. Deep learning methods preserve fine detail better than traditional interpolation.

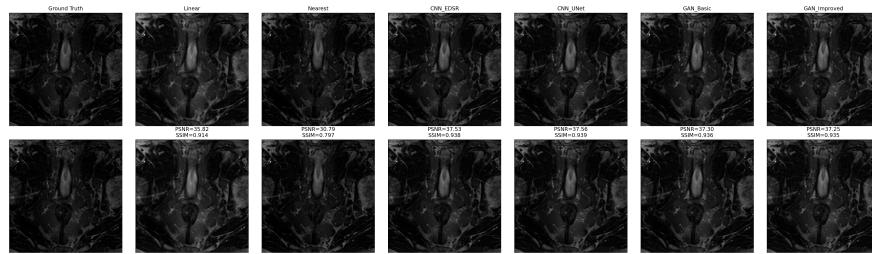


Fig. 2: Side-by-side comparison across two anatomical levels showing consistency of method performance. All deep learning methods generalize well across different anatomical regions.

### 3.2 Additional Qualitative Examples

Figure 1 provides zoomed views highlighting preservation of fine anatomical structures. U-Net and CNN methods maintain sharp boundaries at the prostatic capsule and gland-seminal vesicle interface. GANs show comparable performance with subtle texture differences. Traditional methods exhibit noticeable blurring.

### 3.3 Failure Cases and Limitations

While deep learning methods substantially outperform traditional interpolation, certain challenging scenarios remain:

**Rapid Anatomical Transitions:** At the prostatic apex and base, where anatomy changes rapidly between slices, all methods show increased error. These regions represent the boundaries of feasible interpolation even with  $k=1$  formulation.

**Low Contrast Regions:** In regions with minimal tissue contrast (e.g., within homogeneous central gland), interpolation accuracy depends heavily on intensity gradients. Subtle structural details may be smoothed.

**Motion Artifacts:** When input slices contain patient motion between acquisitions, interpolation performance degrades as the assumption of anatomical continuity is violated.

**Edge Effects:** At the very top and bottom of the prostate volume, where boundary slices are used, interpolation is inherently more difficult due to asymmetric context.

## 4 Extended DDPM Analysis

### 4.1 Training Dynamics

Despite DDPM’s poor final performance, training proceeded stably:

#### Training Loss Evolution:

- Epoch 1: Train loss 0.1119, Val loss 0.0927
- Epoch 5: Train loss 0.0956, Val loss 0.0861
- Epoch 10: Train loss 0.0891, Val loss 0.0844
- Epoch 16: Train loss 0.0808, Val loss 0.0813 (best)
- Epoch 20: Train loss 0.0798, Val loss 0.0821

Loss decreased steadily with no overfitting (training and validation losses tracked closely). This indicates the model successfully learned the noise prediction task. The failure stems from task mismatch, not training issues.

### 4.2 Why Standard Diffusion Objectives Fail for This Task

DDPM optimizes for:

$$\mathcal{L} = \mathbb{E}_{t, x_0, \epsilon} [||\epsilon - \epsilon_\theta(x_t, t, c)||^2] \quad (1)$$

This objective trains the model to predict noise, enabling stochastic sampling. However, slice interpolation requires:

- **Deterministic output:** Single ground truth slice must be reconstructed
- **Pixel-accurate matching:** PSNR/SSIM metrics penalize any deviation
- **Anatomical plausibility:** Generated structure must match actual anatomy

The stochastic sampling process introduces variability that degrades these metrics even if samples are perceptually reasonable.

### 4.3 Alternative Approaches

Several modifications might improve DDPM performance:

**DDIM Sampling:** Denoising Diffusion Implicit Models (DDIM) [8] enable deterministic sampling, potentially improving reconstruction metrics while maintaining diffusion model benefits.

**Increased Training:** Standard DDPM implementations train for 500-1000 epochs with  $T = 1000$  timesteps. My 20 epochs and  $T = 100$  may have been insufficient.

**Hybrid Objectives:** Combining noise prediction with direct reconstruction loss might bridge the gap between generation and reconstruction objectives.

**Conditional Guidance:** Stronger conditioning on anatomical constraints (e.g., segmentation maps, structural priors) could constrain generation to anatomically valid outputs.

However, given the strong performance and 100 $\times$  faster inference of deterministic models, these modifications appear academically interesting but practically unjustified for this specific task.

Figure 3 provides visual evidence of DDPM’s reconstruction failure. While the model successfully learned the noise prediction task (as evidenced by decreasing training loss), the stochastic sampling process produces outputs with severe anatomical inaccuracies. The error map shows concentrated failures (bright regions) at tissue boundaries and within the prostate gland, where precise reconstruction is most critical. Notably, the DDPM output appears somewhat smoother than linear interpolation but with incorrect anatomical details—a perceptually plausible but quantitatively inaccurate result that highlights the mismatch between generative and reconstructive objectives.

#### 4.4 Qualitative DDPM Failure Analysis

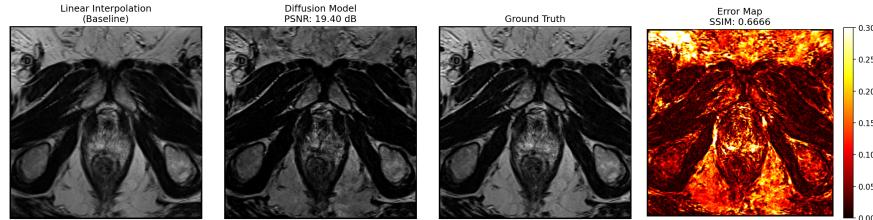


Fig. 3: Visual comparison of DDPM reconstruction failure. Left to right: Linear interpolation baseline, DDPM output (PSNR 19.40 dB), ground truth, error map (SSIM 0.6666). The error map reveals severe reconstruction failures (bright yellow/white regions) across anatomical boundaries, particularly in the prostate gland and surrounding structures. Despite stable training, the stochastic sampling process introduces artifacts that prevent accurate deterministic reconstruction.

## 5 Limitations

**Single Modality:** I focused exclusively on T2-weighted prostate MRI. Generalization to other sequences (T1, DWI), anatomical regions, or imaging modalities requires validation.

**Single Vendor/Field Strength:** All data came from 3T Siemens scanners. Performance on other vendors (GE, Philips) or field strengths (1.5T, 7T) is unknown.

**Resolution Assumptions:** My approach assumes consistent 1.5mm slice spacing. Clinical protocols vary (0.75-3mm), and thicker slices might exceed interpolation feasibility even with  $k=1$  formulation.

**Evaluation Metrics:** PSNR and SSIM are imperfect proxies for clinical utility. Future work should include radiologist evaluation for diagnostic tasks (tumor detection, staging).

**Computational Resources:** My work required substantial GPU resources (hundreds of GPU-hours). This may limit reproducibility in resource-constrained settings, though inference is fast once models are trained.

**Patient Demographics:** The UCLA dataset may not represent all patient populations. Validation on diverse cohorts would strengthen generalizability claims.

## 6 Clinical Implications and Future Directions

### 6.1 Clinical Potential

Improved through-plane resolution enables:

**3D Reconstruction:** Higher isotropy improves 3D rendering for surgical planning and patient education, reducing interpolation artifacts in reconstructed volumes.

**Multiplanar Reformatting:** Clinicians often reformat axial acquisitions to sagittal or coronal views. Better through-plane resolution reduces reformatting artifacts and improves diagnostic confidence in non-native planes.

**Computer-Aided Diagnosis:** Many CAD systems assume isotropic data. Interpolation could enable deployment of such systems on clinical data without requiring acquisition protocol changes.

**Radiation Therapy Planning:** Improved resolution assists in precise tumor delineation for treatment planning.

### 6.2 Deployment Considerations

Clinical deployment requires:

**Regulatory Approval:** FDA clearance or equivalent for algorithm changes to clinical workflow.

**Clear Labeling:** Generated intermediate slices must be clearly marked to prevent misinterpretation as acquired data.

**Validation Studies:** Radiologist reader studies comparing diagnostic performance on original versus interpolated volumes.

**Integration:** Seamless integration into clinical PACS systems and radiology workflows.

### 6.3 Future Research Directions

**Multi-Sequence Integration:** Prostate MRI typically acquires multiple sequences (T2, T1, DWI, DCE). Leveraging information across sequences could improve interpolation quality, as different contrasts reveal complementary anatomical information.

**Uncertainty Quantification:** Providing confidence estimates for interpolated slices would help radiologists identify regions requiring scrutiny. Bayesian deep learning or ensemble methods could quantify prediction uncertainty.

**Active Learning:** In scenarios with limited high-resolution training data, active learning could identify most informative samples for annotation, reducing labeling burden.

**Task-Specific Optimization:** Rather than optimizing pixel-level metrics, training could target downstream tasks (tumor detection, segmentation) to ensure interpolation preserves clinically relevant information.

**Generalization Studies:** Evaluating performance across different scanners, institutions, and patient populations would assess real-world deployability.

**Real-Time Application:** Exploring deployment scenarios where interpolation occurs during or immediately after acquisition to provide higher resolution to radiologists without protocol changes.

## References

- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of Digital Imaging* **26**(6), 1045–1057 (2013). <https://doi.org/10.1007/s10278-013-9622-7>, <https://doi.org/10.1007/s10278-013-9622-7>
- Globus Team: Globus: Research data management. <https://www.globus.org> (2024), accessed: 2024-12-05
- Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851 (2020), <https://arxiv.org/abs/2006.11239>
- Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134 (2017). <https://doi.org/10.1109/CVPR.2017.632>, <https://arxiv.org/abs/1611.07004>
- Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 136–144 (2017), [https://openaccess.thecvf.com/content\\_cvpr\\_2017\\_workshops/w12/papers/Lim\\_Enhanced\\_Deep\\_Residual\\_CVPR\\_2017\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf), arXiv:1707.02921
- Natarajan, S., Priester, A., Margolis, D., Huang, J., Marks, L.S.: Prostate mri and ultrasound with pathology and coordinates of tracked biopsy (prostate-mri-us-biopsy). The Cancer Imaging Archive (2020). <https://doi.org/10.7937/TCIA.2020.A61IOC1A>, <https://www.cancerimagingarchive.net/collection/prostate-mri-us-biopsy/>

7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28), <https://arxiv.org/abs/1505.04597>
8. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021), <https://arxiv.org/abs/2010.02502>
9. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. In: International Conference on Machine Learning. pp. 7354–7363. PMLR (2019), <https://arxiv.org/abs/1805.08318>