

Bayesian Image Reconstruction using Deep Generative Models

Răzvan V. Marinescu

Massachusetts Institute of Technology



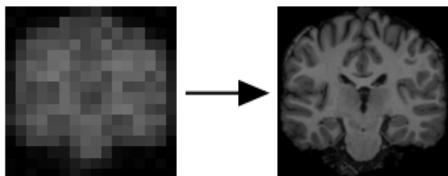
Aim: image reconstruction using ***pre-trained*** generator models

StyleGAN2 (Karras et al, 2019)



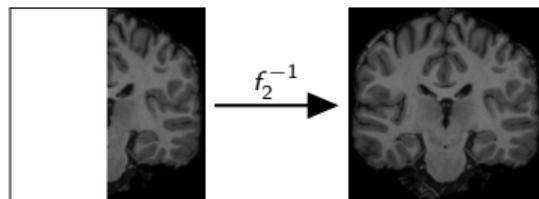
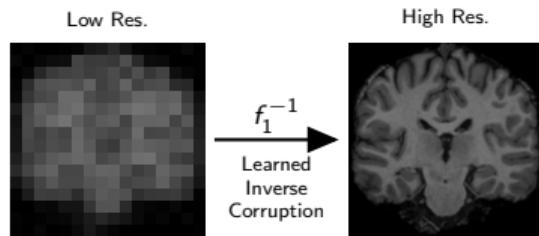
- Adapt the state-of-the-art StyleGAN2 for medical image reconstruction

MRI reconstruction



Current image reconstruction methods have several limitations

- ▶ Require large computational resources and data
- ▶ Are specific to particular corruption tasks
- ▶ Cannot deal with distribution shifts:
 - ▶ in inputs: e.g. older populations
 - ▶ in corruption type: e.g. change in blur kernel



Limitation 1: State-of-the-art DL methods have large computational requirements

- Requirements = Computation Time + Advanced Hardware + Large Datasets
- Most computation now runs on clouds

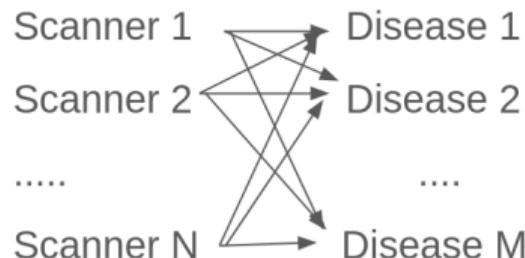


- Currently few labs/companies have the resources to train state-of-the-art models
 - StyleGAN2: 9 days on 4 GPUs
 - GPT-3: 355 years on single GPU
- Solutions moving forward:
 - Adapting previously-trained models
 - Combine smaller models into larger ones

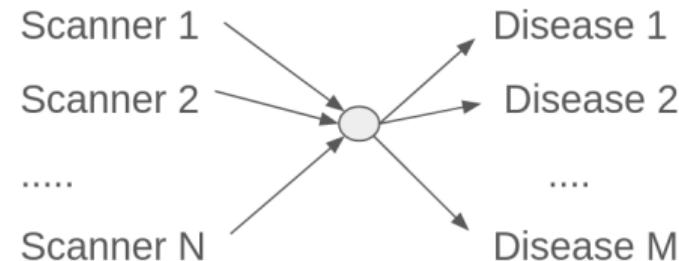
Limitation 2: Distribution shifts require model re-training

- ▶ Distribution shifts happen all the time:
 - ▶ Changes in hospital scanners, protocols, software upgrades
 - ▶ Can be continuous: population getting older due to better healthcare
- ▶ Shifts can result in combinatorial effects in number of re-training instances!
- ▶ Compositionality is one potential solution

Without compositionality: **N x M**



With compositionality: **N + M**



Recent models can perform image reconstruction using **pre-trained** generative models

Image2StyleGAN++ (Abdal et al, 2020)



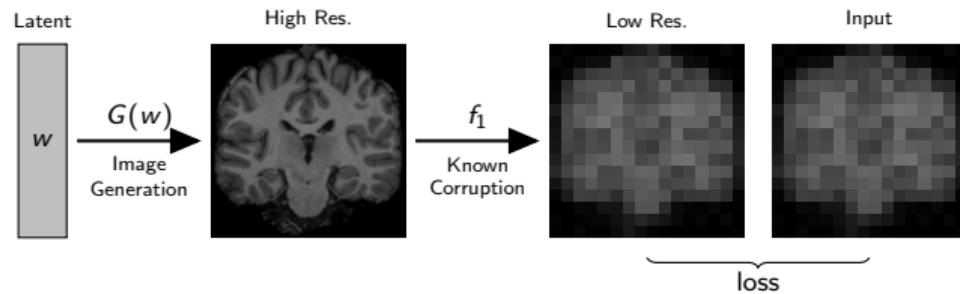
PULSE (Menon et al, 2020)



- ▶ Don't use an embedder network, so can generalise beyond the training data
- ▶ However, they are specific to a particular corruption task (either in-painting or super-resolution)
- ▶ **Cannot characterize uncertainty and recover multiple solutions**
- ▶ We will aim to construct a Bayesian formulation that can fully characterize the posterior over all potential solutions

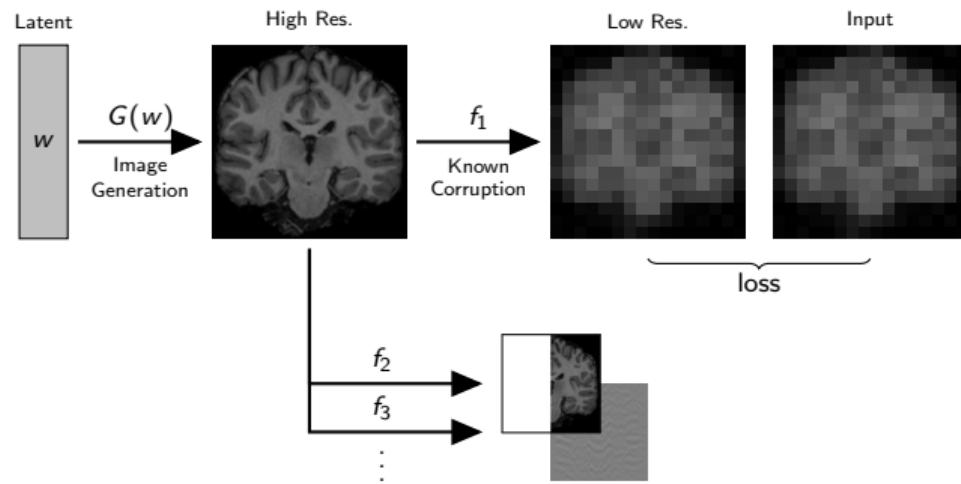
Method: We perform image reconstruction by combining two models

1. a pre-trained generator G (StyleGAN2)
2. a known forward corruption model f_1



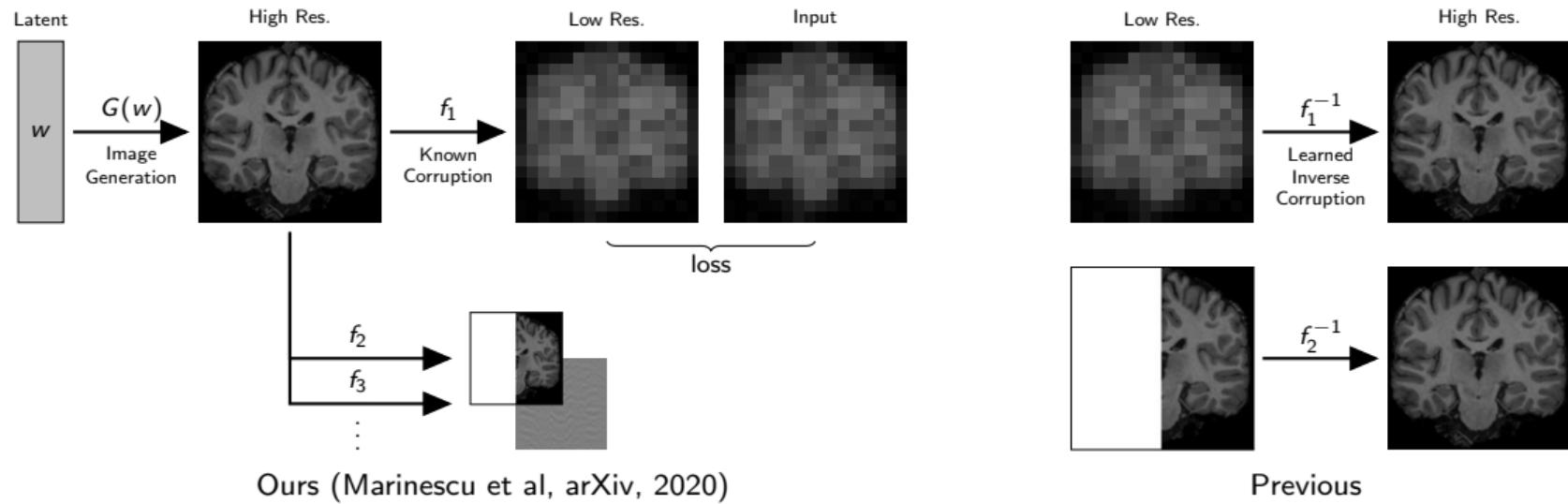
Method: We perform image reconstruction by combining two models

1. a pre-trained generator G (StyleGAN2)
2. a known forward corruption model f_1



Method: We perform image reconstruction by combining two models

1. a pre-trained generator G (StyleGAN2)
2. a known forward corruption model f_1



Reconstructed image is given by computing the Bayesian maximum a-posteriori (MAP) estimate

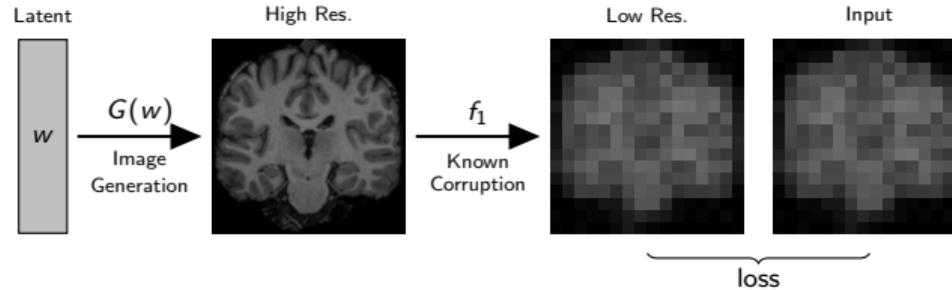
- We optimise:

$$w^* = \arg \max_w p(w)p(I|w)$$

- For uninformative prior $p(w)$ and Gaussian noise model (pixelwise independent), we get:

$$w^* = \arg \min_w \|I - f \circ G(w)\|_2^2$$

- This can be optimised with SGD
- Once we get w^* , the the reconstructed image is $G(w^*)$



Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion

Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes

$$w^*, \eta^* = \arg \min_{w, \eta} \|\phi(I) - \phi \circ f \circ G(w, \eta)\|_2^2$$



Good reconstructions require further modifications

- ▶ We started from the original StyleGAN2 inversion
 - ▶ Yet the reconstruction was not good → required several changes
 - ▶ remove noise layers

$$w^* = \arg \min_w ||\phi(I) - \phi \circ f \circ G(w)||_2^2$$

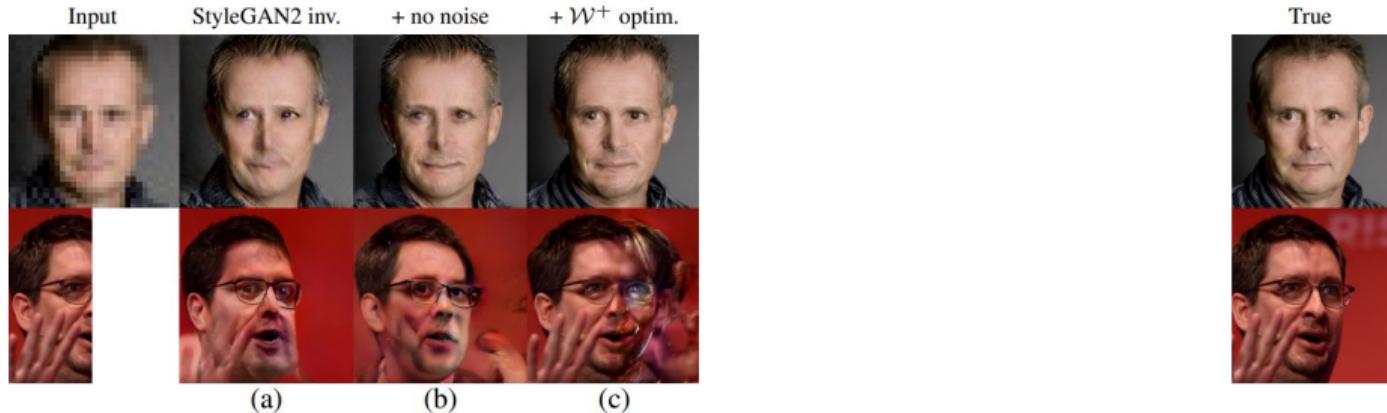


Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
 - optimize latents at all resolutions

$$\mathbf{w} = \mathbf{w}_1, \dots, \mathbf{w}_L$$

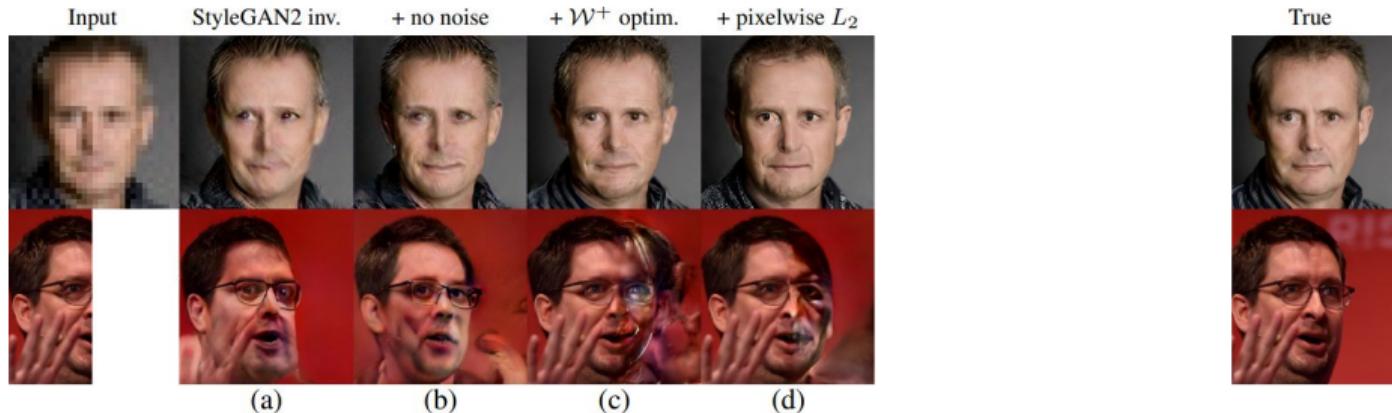
$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\phi(I) - \phi \circ f \circ G(\mathbf{w})\|_2^2$$



Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
 - add pixelwise loss

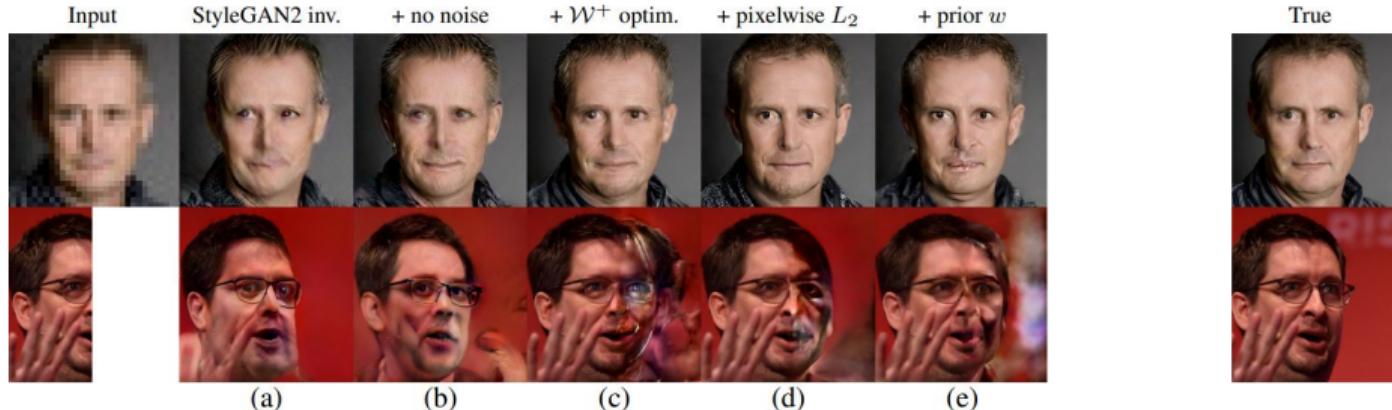
$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\phi(I) - \phi \circ f \circ G(\mathbf{w})\|_2^2 + \|I - f \circ G(\mathbf{w})\|_2^2$$



Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
 - gaussian prior on latents

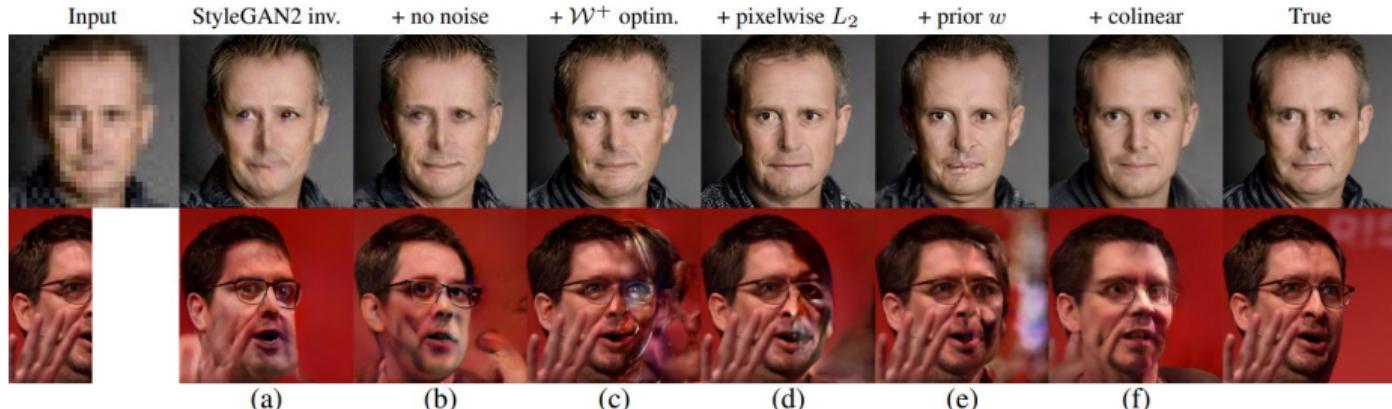
$$\begin{aligned}\mathbf{w}^* = \arg \min_{\mathbf{w}} & ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2 + ||I - f \circ G(\mathbf{w})||_2^2 + \\ & + \sum_i \left(\frac{w_i - \mu}{\sigma_i} \right)^2\end{aligned}$$



Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
 - force latents to be colinear

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\phi(I) - \phi \circ f \circ G(\mathbf{w})\|_2^2 + \|I - f \circ G(\mathbf{w})\|_2^2 + \\ + \sum_i \left(\frac{\mathbf{w}_i - \mu}{\sigma_i} \right)^2 - \sum_{i,j} \frac{\mathbf{w}_i \mathbf{w}_j^T}{\|\mathbf{w}_i\| \|\mathbf{w}_j\|}$$



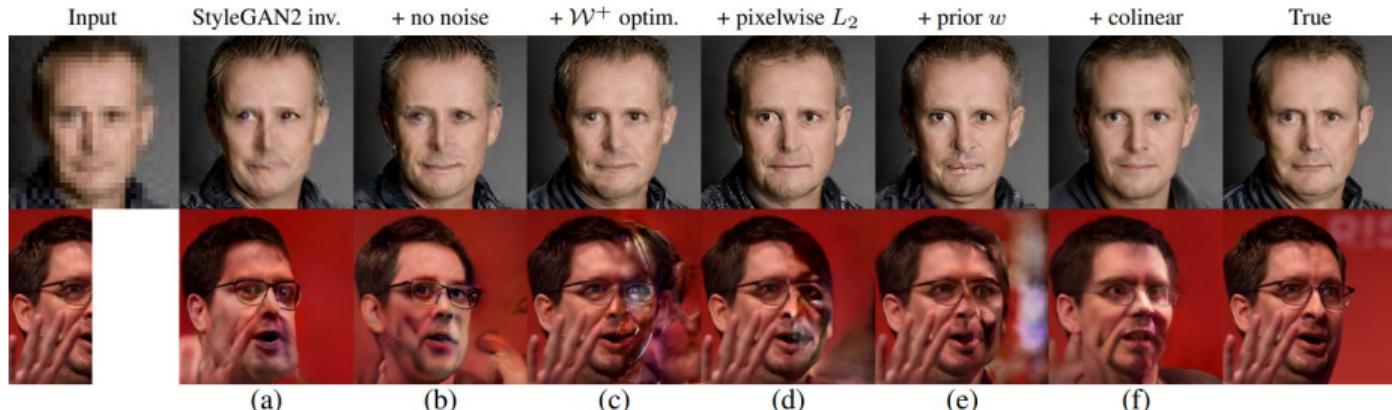
Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion

- Yet the reconstruction was not good → required several changes

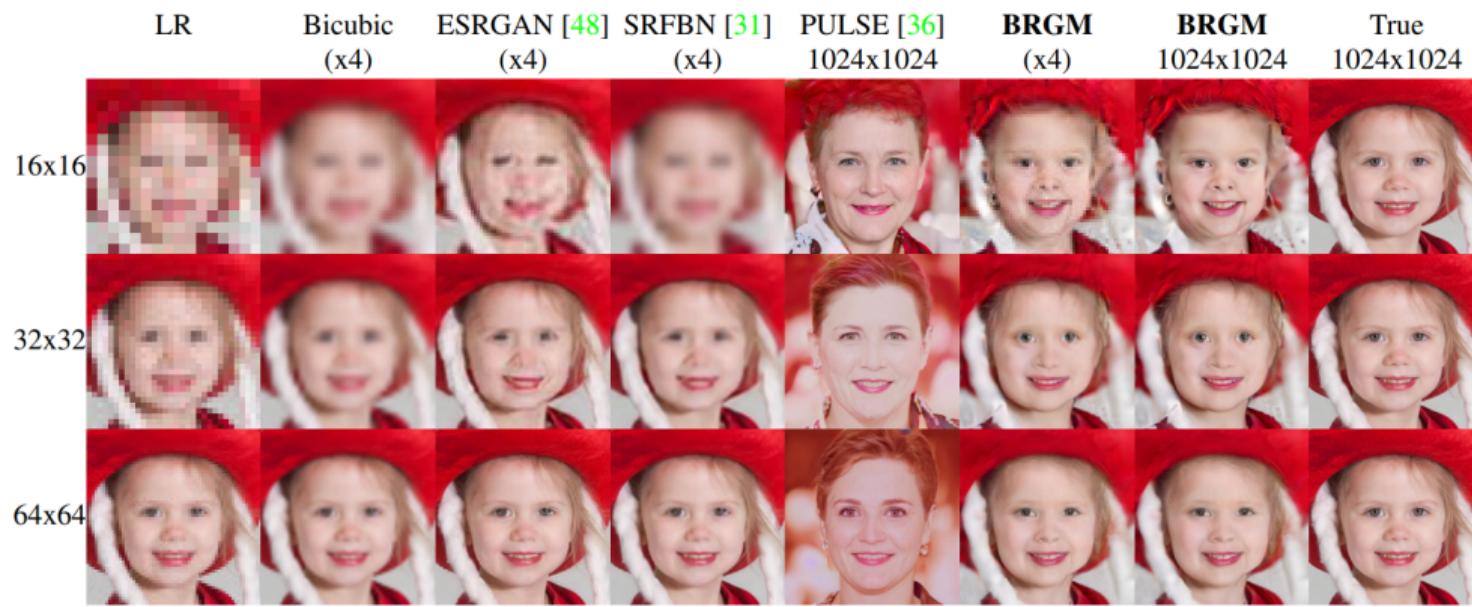
- Analytically expressed the full likelihood (Marinescu et al, 2021)

$$\begin{aligned} \mathbf{w}^* = \arg \min_{\mathbf{w}} & ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2 + ||I - f \circ G(\mathbf{w})||_2^2 + \\ & + \sum_i \left(\frac{\mathbf{w}_i - \mu}{\sigma_i} \right)^2 - \sum_{i,j} \frac{\mathbf{w}_i \mathbf{w}_j^T}{|\mathbf{w}_i| |\mathbf{w}_j|} \end{aligned}$$



Results on super-resolution using the FFHQ dataset

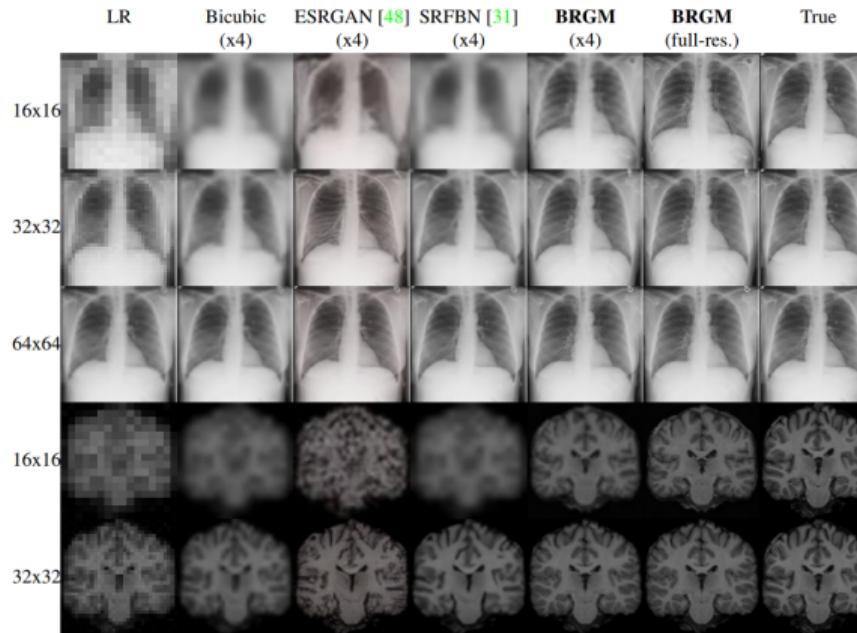
- We achieve state-of-the-art (SOTA) results on small inputs resolutions 16x16
- On larger resolutions (>32x32), we achieve very good results, albeit not SOTA



Marinescu et al, arXiv, 2020

Similar results on super-resolution for medical datasets

- We achieve state-of-the-art (SOTA) results on small inputs resolutions 16x16
- On larger resolutions (>32x32), we achieve very good results, albeit not SOTA



Marinescu et al, arXiv, 2020

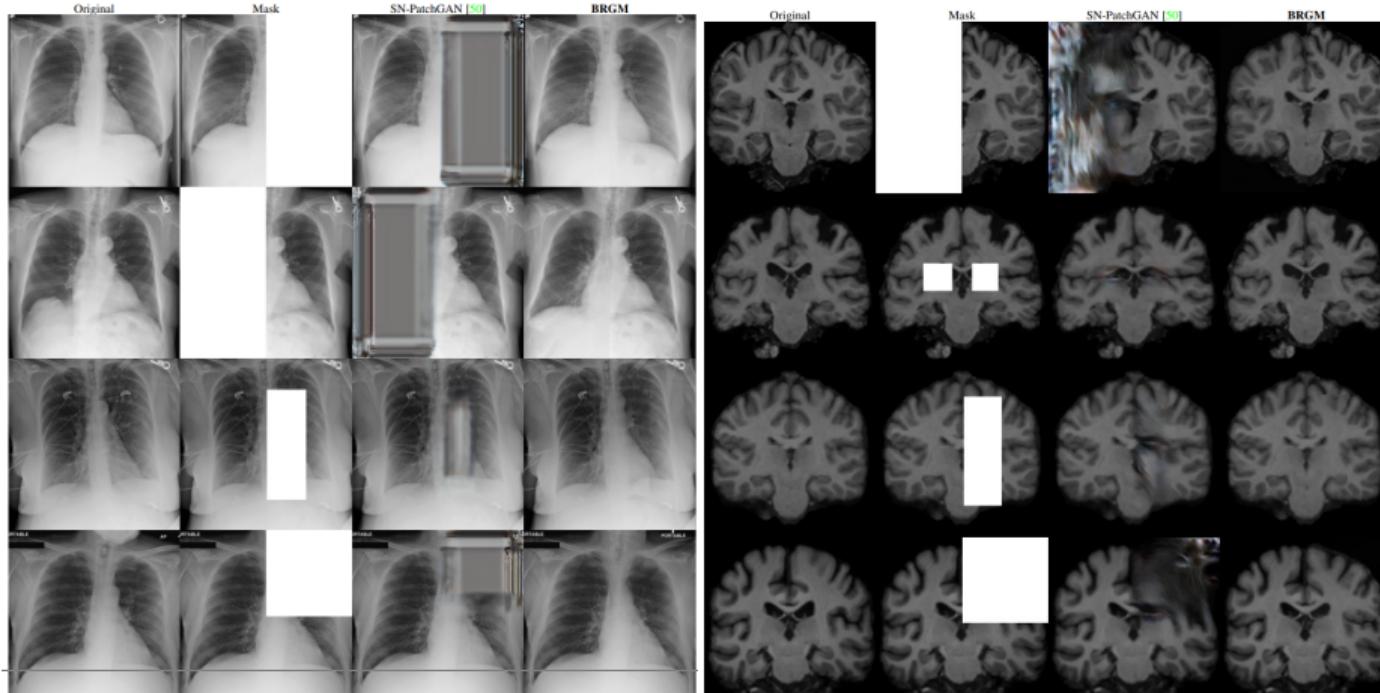
Inpainting also achieves state-of-the-art results

- Best previous method (SN-PatchGAN, CVPR 2019) does not work for large masks
- Our method can “hypothesize” missing structure



Inpainting also achieves state-of-the-art results

- Best previous method (SN-PatchGAN, CVPR 2019) does not work for large masks
- Our method can “hypothesize” missing structure



Results confirmed through quantitative evaluation

- ▶ Three different datasets, at different resolutions
- ▶ Human study with 20 raters

Super-resolution

Dataset	BRGM	PULSE [36]	ESRGAN [48]	SRFBN [31]
FFHQ 16 ²	0.24 /25.66	0.29/27.14	0.35/29.32	0.33/ 22.07
FFHQ 32 ²	0.30/18.93	0.48/42.97	0.29/23.02	0.23 / 12.73
FFHQ 64 ²	0.36/16.07	0.53/41.31	0.26/18.37	0.23 / 9.40
FFHQ 128 ²	0.34/15.84	0.57/34.89	0.15/15.84	0.09 / 7.55
X-ray 16 ²	0.18 / 11.61	-	0.32/14.67	0.37/12.28
X-ray 32 ²	0.23/10.47	-	0.32/12.56	0.21 / 6.84
X-ray 64 ²	0.31/10.58	-	0.30/8.67	0.22 / 5.32
X-ray 128 ²	0.27/10.53	-	0.20/7.19	0.07 / 4.33
Brains 16 ²	0.12 / 12.42	-	0.34/22.81	0.33/12.57
Brains 32 ²	0.17 /11.08	-	0.31/14.16	0.18/ 6.80

Inpainting

Dataset	BRGM				SN-PatchGAN [50]			
	LPIPS	RMSE	PSNR	SSIM	LPIPS	RMSE	PSNR	SSIM
FFHQ	0.19	24.28	21.33	0.84	0.24	30.75	19.67	0.82
X-ray	0.13	13.55	27.47	0.91	0.20	27.80	22.02	0.86
Brains	0.09	8.65	30.94	0.88	0.22	24.74	21.47	0.75

Human evaluation

Dataset	BRGM	PULSE [36]	ESRGAN [48]	SRFBN [31]
FFHQ 16 ²	0.42	0.32	0.11	0.15
FFHQ 32 ²	0.39	0.02	0.12	0.47
FFHQ 64 ²	0.14	0.08	0.32	0.45
FFHQ 128 ²	0.14	0.10	0.39	0.38

Our method also has limitations that we plan to address

- ▶ It can fail for images that are too dissimilar to the training ones
 - ▶ Because generator cannot extrapolate easily



- ▶ Can be inconsistent with the input image



- ▶ Proposed a method for image reconstruction using pre-trained deep generative models
- ▶ Solution is given by the Bayesian MAP estimate
- ▶ State-of-the-art results on super-resolution and inpainting