Self-supervised learning for medical image analysis using image context restoration

Liang Chen[a,b,*], Paul Bentley[b], Kensaku Mori[c], Kazunari Misawa[d], Michitaka Fujiwara[e], Daniel Rueckert[a]

a BioMedIA Group, Department of Computing, Imperial College London, 180 Queen's Gate, London, SW7 2AZ, UK

b Division of Brain Sciences, Department of Medicine, Imperial College London, UK

c Graduate School of Informatics, Nagoya University, Japan

d Aichi Cancer Centre, Japan

e Nagoya University Hospital, Japan

Presented by: Xi Fang

Date: 09/04/2019

Content

- Motivation
- Contribution
- Related Works
- Methods
- Experiments
- Results
- Conclusion

Motivation

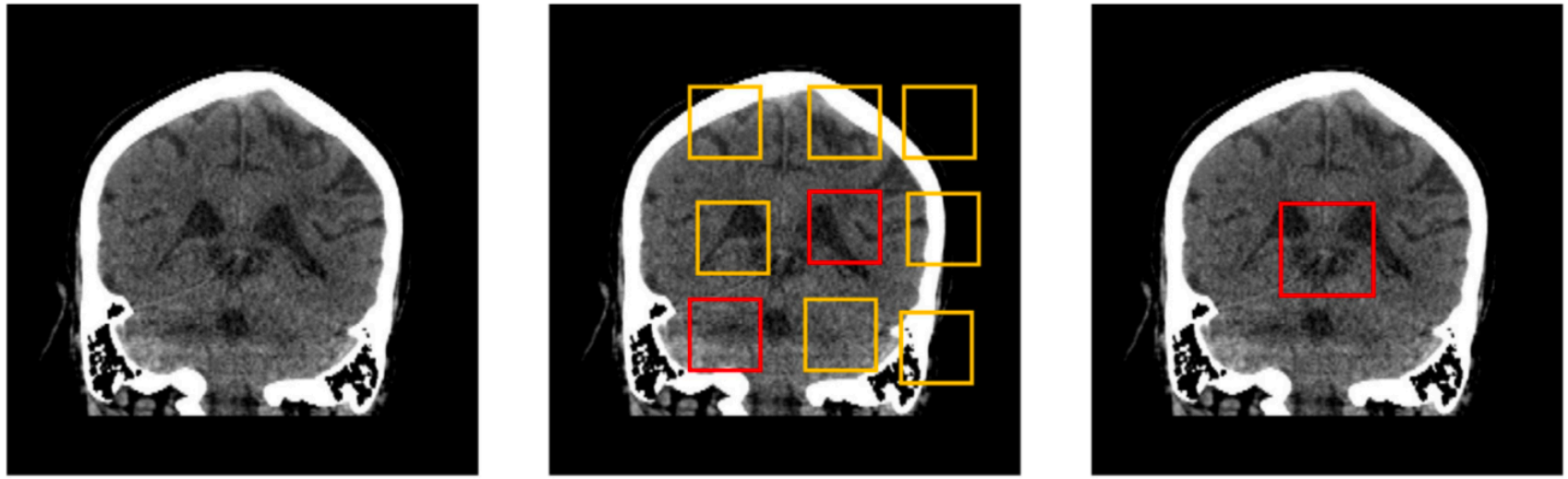
- CNNs have demonstrated significant improvement when applied to challenging tasks such as disease classification and organ segmentation
- Training CNNs only using the small number of labelled images cannot always achieve satisfactory results and does not exploit the potentially large number of unlabeled images that may be available.
- The pretrained models from the natural image domain are not useful in the medical imaging domain since the intensity distribution of natural images is different from that of medical images

Contribution

- Propose a novel self-supervised learning strategy based on context restoration to initialize model training.
- It learns semantic features.
- These image features are useful for different types of subsequent image analysis tasks.
- Its implementation is simple.
- It learns useful semantic features and lead to improved machine learning models for different tasks.

Related Works

- Autoencoder + L2 Loss
- Prediction of the relative positions of image patches (RP method)
- Local context prediction (CP method)

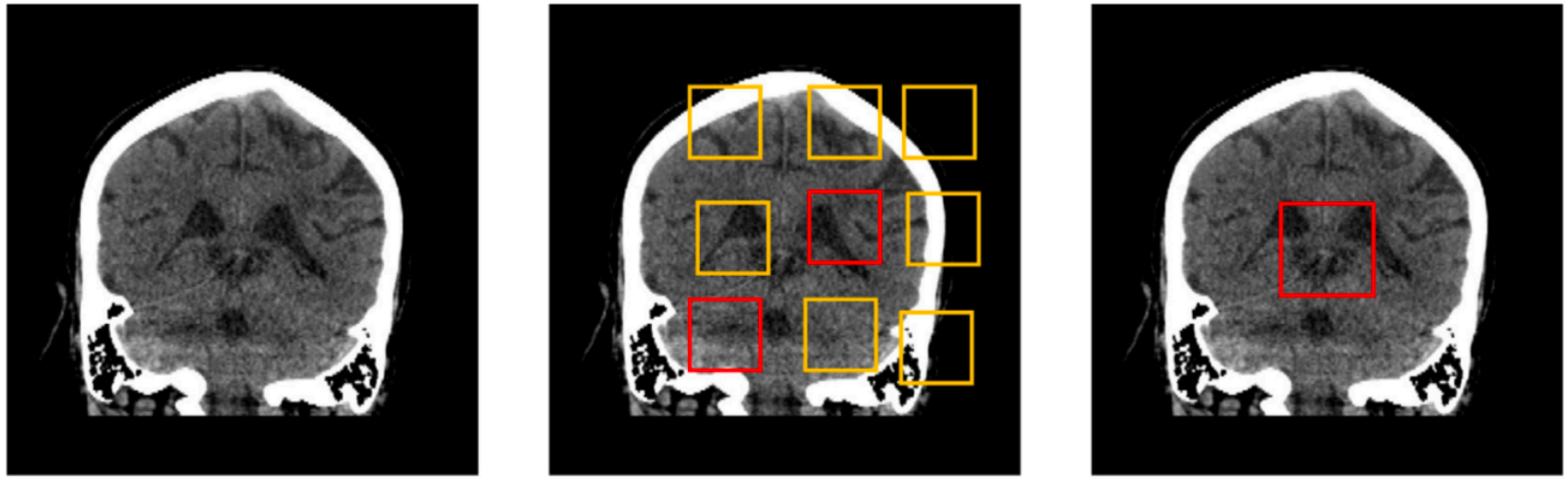


(a) Original image (b) RP method (c) CP method

Fig. 1. Demonstration of the RP and CP method on a brain CT image. (a) shows the original CT image in the coronal view. (b) shows the patch grid of the RP method and the red rectangles indicate patches of left cerebellum and right cerebrum. (c) shows the selected patch to be predicted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Related Works

- Autoencoder + L2 Loss Features that have limited value for discriminative tasks
- Prediction of the relative positions of image patches (RP method) No global context of images
- Local context prediction (CP method) reconstruct image-level maps the removal of context changes the image intensity distribution

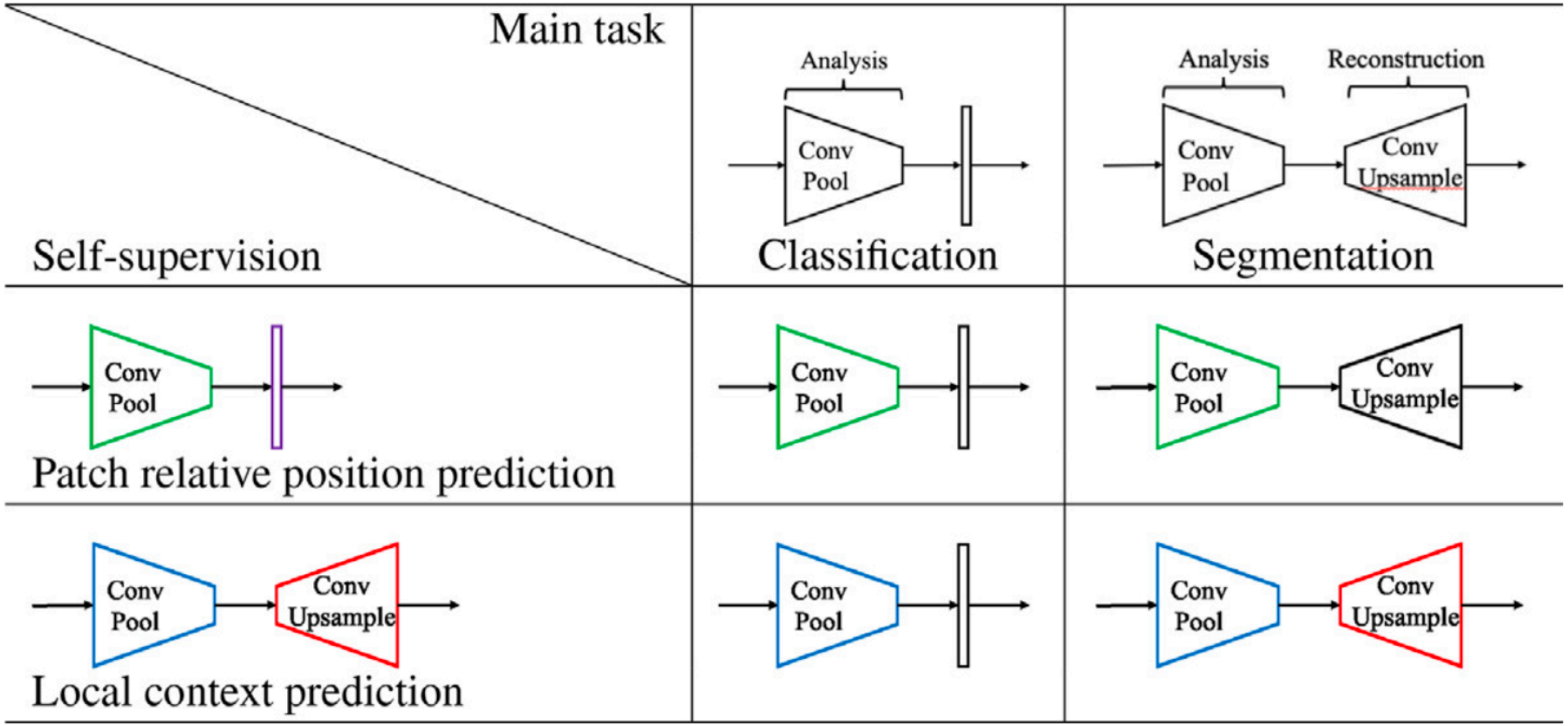


(a) Original image (b) RP method (c) CP method

Fig. 1. Demonstration of the RP and CP method on a brain CT image. (a) shows the original CT image in the coronal view. (b) shows the patch grid of the RP method and the red rectangles indicate patches of left cerebellum and right cerebrum. (c) shows the selected patch to be predicted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Related Works

Comparison between the RP method and the CP method. Weights learned in both of them can ir
tialise the subsequent classification CNN. Weights learned in the RP method can only initialise t
analysis part of the subsequent segmentation CNN; while weights learning in the CP method can ir
tialise analysis and reconstruction part of the subsequent segmentation CNN.



Related Works

- Jigsaw (JS) method:
  - a more complicated version of patch relative positions, in which all 9 patches are input to CNNs in a random sequence. The CNNs were trained to find the correct sequence of the patches.

Methods

- *Context restoration*
  - Given a dataset X = {**x**1 , **x**2 , . . . , **x***N* } consisting of *N* images with no annotations, a new dataset X̃ = *f* (X) is generated.
  - Here X̃ = {**x̃**1,**x̃**2,...,**x̃***N*}.*f*(·) is a function corrupting the context of original images.
  - **x***i* = *g*(**x̃***i*) = *f*−1(**x̃***i*),

Algorithm 1: Image context disordering.

Input: original image $\mathbf{x_i}$
Output: image with disordered context $\tilde{\mathbf{x}}_\mathbf{i}$
for $iter = 1, 2, \ldots, T$ **do**
  randomly select a patch $p_1 \in \mathbf{x_i}$
  randomly select a patch $p_2 \in \mathbf{x_i}$
  $p_1 \cap p_2 = \emptyset$
  swap p_1 and p_2

Methods

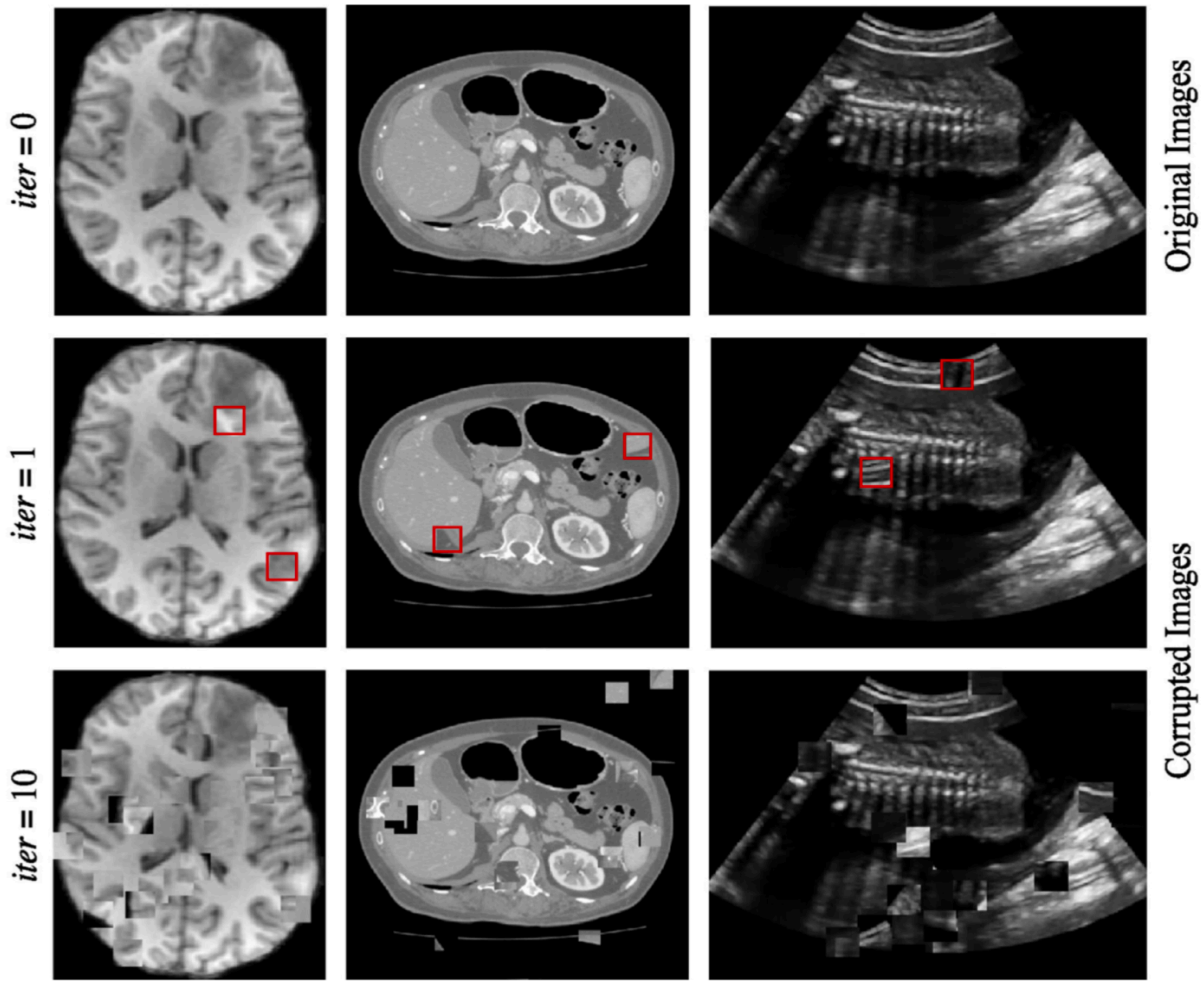


Fig. 2. Generating training images for self-supervised context disordering: Brain T1 MR image, abdominal CT image, and 2D fetal ultrasound image, respectively. In figures in the second column, red boxes highlight the swapped patches after the first iteration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Methods

- Network architectures

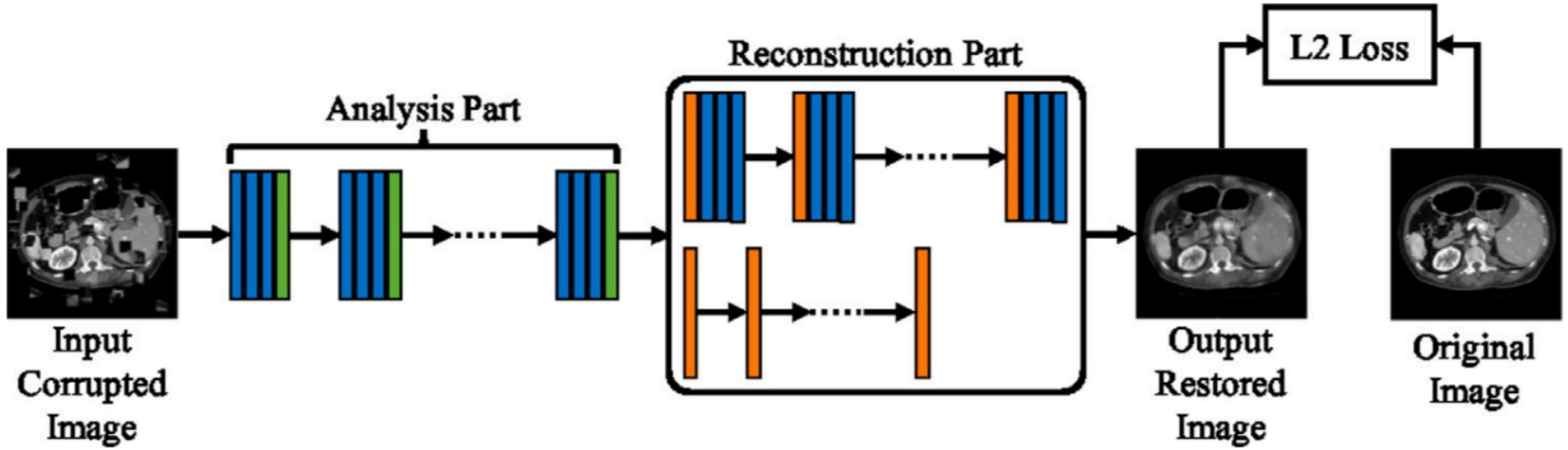


Fig. 3. General CNN architecture for the context restoration self-supervised learning. In the figure, the blue, green, and orange strides represent convolutional units, down-sampling units, and upsampling units, respectively. In the reconstruction part, CNN structures could vary depending on subsequent task type. For subsequent classification tasks, the simple structures such as a few deconvolution layers (2nd row) are preferred. For subsequent segmentation tasks, the complex structures (1st row) consistent with the segmentation CNNs are preferred. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Methods

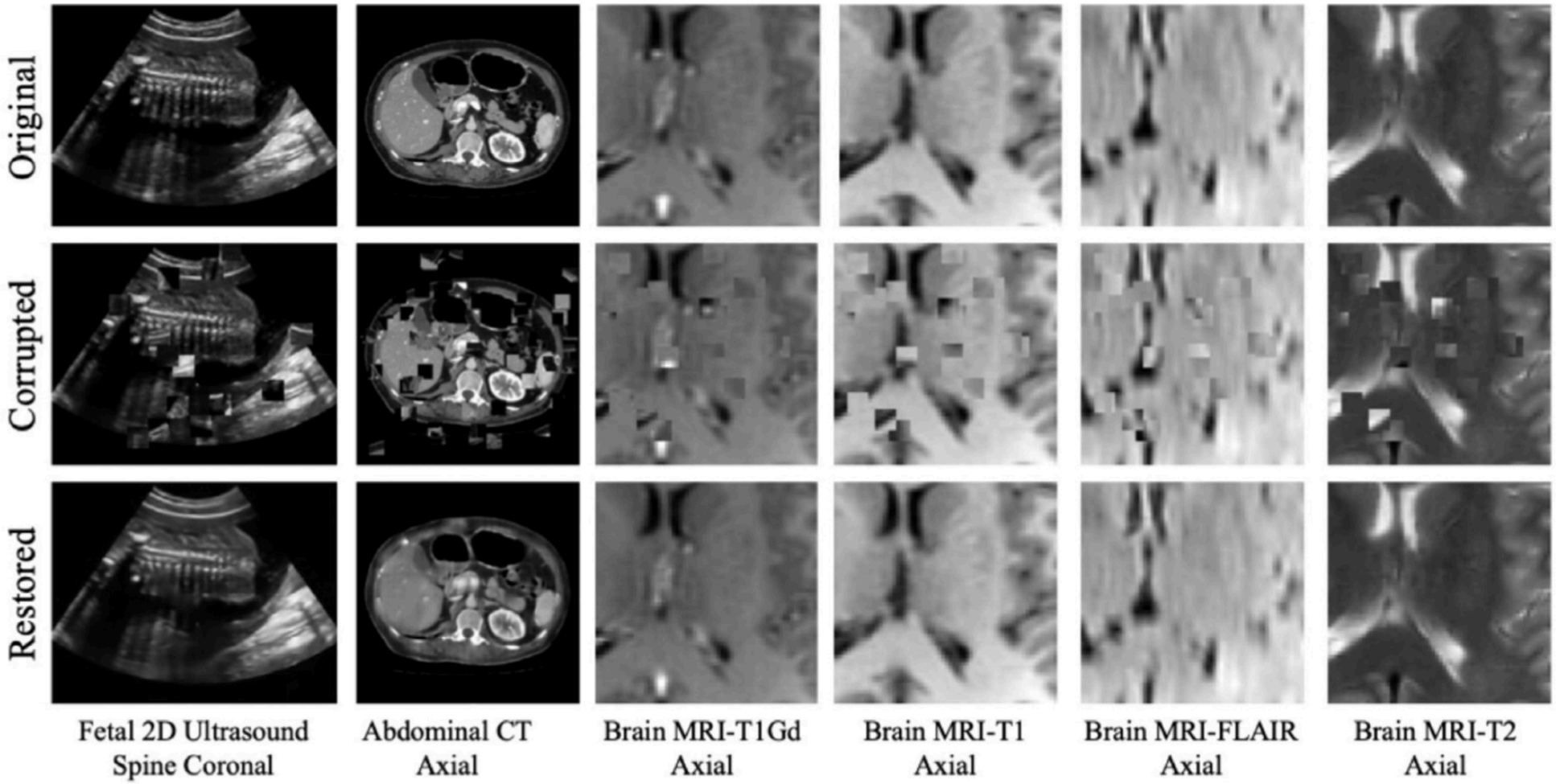


Fig. 4. Self-supervision using context restoration: For brain MR images, our training is on 2D image patch level. Therefore, the context restoration is also based on patches.

Experiments

- Classification:
  - Datasets:
    - 2694 2D ultrasound examinations of fetuses with gestational ages between 18 and 22 weeks
  - Implementation:
    - The CNN for this classification problem is the SonoNet-64
- Localization:
  - Datasets:
    - A dataset of 3D abdominal CT image from 150 subjects is employed
  - Implementation:
    - The CNN is similar to SonoNet but has one more new stack of convolution and pooling layers

Experiments

- Segmentation
  - Datasets:
    - The dataset of the BraTS 2017 challenge which consists of 285 subjects
  - Implementation:
    - A 2D patch-based CNN approach
    - The patch size used is 64×64
    - The CNN used in this ex- periment is a 2D U-Net

Results

- Classification

Table 3
The classification of standard scan planes of fetal 2D ultrasound images. The entries in bold highlight the best comparable results.

Training	Initialisation	Precision (%)	Recall (%)	F1-score (%)
100% (Baumgartner et al., 2017)	Random	80.60	86.00	82.80
100%, Ours	Random	**89.39**	**89.66**	**89.42**
50%	Random	84.69	84.94	84.64
	Random + augmentation	84.09	84.86	84.06
	Auto-encoder (Bengio et al., 2007)	84.63	86.09	84.50
	Relative positions (Doersch et al., 2015)	85.15	86.79	84.74
	Jigsaw (Noroozi and Favaro, 2016)	84.89	86.96	85.4
	Context prediction (Pathak et al., 2016)	84.43	85.27	84.43
	Context restoration	**85.52**	**87.56**	**85.94**
25%	Random	57.23	78.99	62.85
	Random + augmentation	60.48	76.23	64.16
	Auto-encoder (Bengio et al., 2007)	55.54	82.87	62.32
	Relative positions (Doersch et al., 2015)	61.01	83.09	66.38
	Jigsaw (Noroozi and Favaro, 2016)	61.56	79.54	65.81
	Context prediction (Pathak et al., 2016)	57.73	81.58	63.10
	Context restoration	**65.69**	**85.25**	**69.93**

Results

- Localization

The performance of the CNN solving the multi-organ localization problem in different training settings. The entries in bold highlight the best comparable results. The RD, AE, RP, JS, CP, CR are short for random, auto-encoder (Bengio et al., 2007), relative positions (Doersch et al., 2015), jigsaw(Noroozi and Favaro, 2016), context prediction (Pathak et al., 2016), and our proposed context restoration. The numbers displayed are the mean ± std distances in mm.

Train	Init.	Left kidney		Right kidney	
		Centroid	Wall	Centroid	Wall
100%	RD	6.45 ± 8.47	3.68 ± 21.41	5.71 ± 10.17	2.79 ± 23.65
50%	RD	17.49 ± 49.67	9.36 ± 75.00	10.40 ± 30.37	5.89 ± 48.28
	RD+AG	67.22 ± 123.47	37.79 ± 180.12	21.82 ± 58.52	11.78 ± 88.17
	AE	12.79 ± 38.67	6.84 ± 56.97	20.44 ± 41.48	11.52 ± 67.01
	RP	12.11 ± 39.01	6.75 ± 61.67	10.61 ± 30.41	5.77 ± 48.64
	JS	14.15 ± 35.71	7.87 ± 55.31	22.31 ± 58.4	12.29 ± 87.93
	CP	11.95 ± 38.97	6.82 ± 61.23	8.30 ± 11.92	4.47 ± 27.83
	CR	**5.99 ± 9.83**	**3.16 ± 22.66**	**5.83 ± 10.10**	**2.90 ± 22.04**
25%	RD	28.23 ± 71.95	15.87 ± 107.18	12.71 ± 30.39	6.77 ± 49.26
	RD+AG	52.19 ± 105.74	29.72 ± 154.70	56.56 ± 98.47	31.62 ± 146.47
	AE	25.90 ± 65.64	14.40 ± 98.28	36.28 ± 73.65	19.55 ± 111.46
	RP	27.65 ± 75.31	15.41 ± 111.82	**8.34 ± 11.22**	**3.97 ± 23.26**
	JS	40.21 ± 89.41	23.30 ± 132.33	66.62 ± 102.08	15.17 ± 43.75
	CP	21.86 ± 60.28	13.03 ± 90.92	15.58 ± 35.3	8.42 ± 57.53
	CR	**7.63 ± 9.02**	**3.94 ± 22.78**	17.51 ± 52.67	9.8 ± 78.57

Results

- Localization

Train	Init.	Pancreas		Liver		Spleen	
		Centroid	Wall	Centroid	Wall	Centroid	Wall
100%	RD	13.39 ± 9.73	8.98 ± 23.27	7.50 ± 5.22	4.35 ± 14.07	6.63 ± 9.68	4.10 ± 23.02
50%	RD	16.45 ± 9.00	10.74 ± 26.77	12.79 ± 8.19	6.89 ± 22.6	13.24 ± 36.97	8.54 ± 56.87
	RD+AG	18.25 ± 11.23	12.75 ± 31.70	13.73 ± 9.28	7.17 ± 24.95	17.86 ± 48.84	10.64 ± 74.25
	AE	15.59 ± 8.51	10.35 ± 24.35	14.07 ± 8.66	7.41 ± 24.39	12.36 ± 11.31	8.54 ± 31.16
	RP	15.54 ± 7.98	11.13 ± 23.50	10.12 ± 8.85	6.18 ± 22.31	7.64 ± 10.16	4.77 ± 24.41
	JS	16.81 ± 9.52	12.00 ± 26.13	16.63 ± 14.37	11.08 ± 40.40	12.78 ± 8.34	7.53 ± 23.22
	CP	14.76 ± 8.78	**10.07 ± 26.26**	9.91 ± 6.78	5.03 ± 15.39	7.79 ± 11.41	4.82 ± 25.98
	CR	**14.76 ± 8.10**	10.14 ± 24.86	**8.91 ± 6.20**	**4.67 ± 16.83**	**7.07 ± 9.54**	**4.05 ± 22.17**
25%	RD	22.09 ± 11.72	17.14 ± 39.23	12.02 ± 6.46	7.14 ± 20.27	24.86 ± 36.64	15.30 ± 61.38
	RD+AG	20.60 ± 18.48	16.40 ± 43.83	19.06 ± 12.48	10.77 ± 36.17	24.78 ± 34.56	13.44 ± 59.24
	AE	17.67 ± 8.40	12.24 ± 25.54	16.79 ± 9.47	9.56 ± 28.30	22.65 ± 47.91	13.95 ± 73.05
	RP	17.84 ± 8.94	**11.74 ± 25.06**	15.59 ± 9.79	9.25 ± 29.74	14.51 ± 38.89	9.95 ± 62.12
	JS	17.91 ± 10.54	11.99 ± 27.74	20.49 ± 13.76	13.12 ± 59.83	20.44 ± 37.19	37.91 ± 152.35
	CP	21.81 ± 11.44	18.59 ± 41.57	11.40 ± 8.69	**6.18 ± 22.50**	10.34 ± 9.92	7.56 ± 27.58
	CR	**16.01 ± 8.46**	11.78 ± 28.79	**11.17 ± 9.03**	7.52 ± 25.68	**8.39 ± 6.28**	**5.82 ± 19.50**

Results

- Brain tumor segmentation

The segmentation results of the customised U-Nets (Ronneberger et al., 2015) in different training settings. The entries in bold highlight the best comparable results. The RD, AE, RP, JS, CP, CR are short for random, auto-encoder (Bengio et al., 2007), relative positions (Doersch et al., 2015), jigsaw (Noroozi and Favaro, 2016), context prediction (Pathak et al., 2016), and our proposed context restoration.

Train	Init.	Dice %			Sensitivity %			Specificity %			Hausdorff95		
		Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.
100%	RD	86.56	77.04	66.31	87.05	77.28	77.62	99.88	99.94	99.95	30.78	25.03	25.74
50%	RD	84.41	75.55	65.11	84.75	77.76	80.20	99.86	99.91	99.94	31.29	25.26	26.81
	RD+AG	82.30	73.17	62.82	**88.46**	77.88	72.67	99.78	99.89	**99.95**	50.98	47.61	42.96
	AE	84.33	71.85	65.07	84.71	74.19	77.38	99.87	99.91	**99.95**	33.36	25.24	24.56
	RP	84.38	75.65	66.73	84.65	77.02	79.48	99.87	**99.92**	**99.95**	36.43	23.15	**20.69**
	JS	83.08	72.02	65.55	80.41	74.44	80.04	99.90	99.93	99.94	41.46	33.46	35.76
	CP	84.54	73.86	66.01	84.59	75.28	79.46	99.86	**99.92**	99.94	33.59	28.59	26.90
	CR	**85.57**	**76.20**	**68.24**	83.83	**78.17**	**80.53**	**99.89**	**99.92**	**99.95**	**26.41**	**20.34**	24.38
25%	RD	81.91	71.22	62.57	84.08	75.68	75.98	99.82	99.89	99.94	36.34	37.21	31.57
	RD+AG	81.02	66.69	60.79	79.64	64.49	66.23	99.87	**99.94**	**99.96**	44.59	34.61	33.59
	AE	83.05	68.92	61.28	83.90	76.52	76.75	99.85	99.86	99.93	33.21	34.9	31.95
	RP	82.38	71.33	61.86	84.23	72.53	75.38	99.83	99.92	99.94	37.83	**31.81**	31.04
	CP	83.19	71.55	62.77	**85.75**	73.68	76.88	99.83	99.91	99.94	36.21	36.45	31.90
	JS	82.09	70.81	62.01	81.59	72.60	68.68	**99.88**	99.92	**99.96**	42.68	36.51	33.69
	CR	**84.27**	**73.43**	**64.12**	85.57	**78.79**	**79.14**	99.85	99.89	99.94	**33.15**	32.18	**30.61**

Results

- Segmentation

Table 5
The segmentation results of the customised U-Nets (Ronneberger et al., 2015) in different training settings. The entries in bold highlight the best comparable results. The RD, AE, RP, JS, CP, CR are short for random, auto-encoder (Bengio et al., 2007), relative positions (Doersch et al., 2015), jigsaw (Noroozi and Favaro, 2016), context prediction (Pathak et al., 2016), and our proposed context restoration.

Train	Init.	Dice %			Sensitivity %			Specificity %			Hausdorff95		
		Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.
100%	RD	86.56	77.04	66.31	87.05	77.28	77.62	99.88	99.94	99.95	30.78	25.03	25.74
50%	RD	84.41	75.55	65.11	84.75	77.76	80.20	99.86	99.91	99.94	31.29	25.26	26.81
	RD+AG	82.30	73.17	62.82	**88.46**	77.88	72.67	99.78	99.89	**99.95**	50.98	47.61	42.96
	AE	84.33	71.85	65.07	84.71	74.19	77.38	99.87	99.91	**99.95**	33.36	25.24	24.56
	RP	84.38	75.65	66.73	84.65	77.02	79.48	99.87	**99.92**	**99.95**	36.43	23.15	**20.69**
	JS	83.08	72.02	65.55	80.41	74.44	80.04	99.90	99.93	99.94	41.46	33.46	35.76
	CP	84.54	73.86	66.01	84.59	75.28	79.46	99.86	**99.92**	99.94	33.59	28.59	26.90
	CR	**85.57**	**76.20**	**68.24**	83.83	**78.17**	**80.53**	**99.89**	**99.92**	**99.95**	**26.41**	**20.34**	24.38
25%	RD	81.91	71.22	62.57	84.08	75.68	75.98	99.82	99.89	99.94	36.34	37.21	31.57
	RD+AG	81.02	66.69	60.79	79.64	64.49	66.23	99.87	**99.94**	**99.96**	44.59	34.61	33.59
	AE	83.05	68.92	61.28	83.90	76.52	76.75	99.85	99.86	99.93	33.21	34.9	31.95
	RP	82.38	71.33	61.86	84.23	72.53	75.38	99.83	99.92	99.94	37.83	**31.81**	31.04
	CP	83.19	71.55	62.77	**85.75**	73.68	76.88	99.83	99.91	99.94	36.21	36.45	31.90
	JS	82.09	70.81	62.01	81.59	72.60	68.68	**99.88**	99.92	**99.96**	42.68	36.51	33.69
	CR	**84.27**	**73.43**	**64.12**	85.57	**78.79**	**79.14**	99.85	99.89	99.94	**33.15**	32.18	**30.61**

Conclusion

- Self-supervised is very helpful to leverage large scale of unlabeled data
- Context based method is the most frequent self-supervised method in medical image domain
- The generated label for self-supervised learning is better to contain semantic features