Boosting segmentation with image-to-image translation

Eugene Vorontsov

Co-authored with: Pavlo Molchanov, Chris Beckham, Wonmin Byeon, Shalini de Mello, Varun Jampani, Ming-Yu Liu, Samuel Kadoury, Jan Kautz

OVERVIEW

Background

Method

Evaluation

Disentangling

Further work

Background

Background

Aim

Minimize # reference segmentations required to train segmentation model.

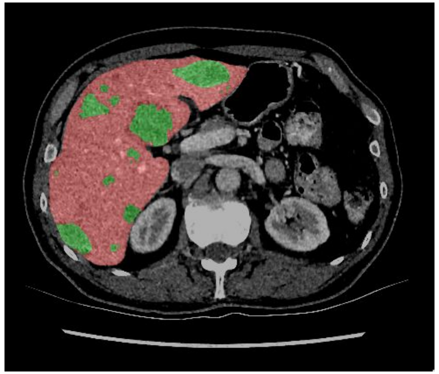
Background

Motivation

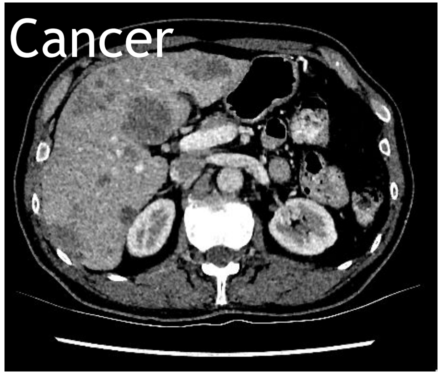
In general : Reduce labeling burden — cheaper, easier, faster.

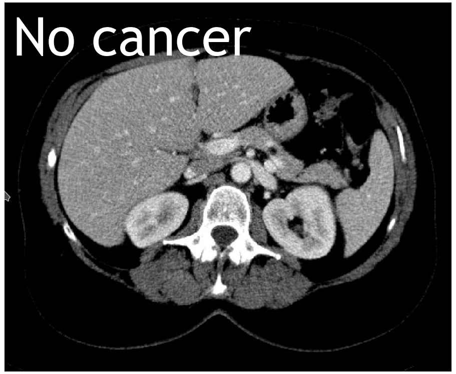
In medical imaging : Hard limit on how quickly images can be segmented.

liver (red); tumor (green)



Few segmentations





Plentiful weak (domain) labels

Background

Related work

Discriminator on segmentation

- (2017 Zhang) In the absence of labels, use discriminator.

Anomaly detection by learning GAN for healthy cases

- (2017 Schlegl, 2018 Baur, 2018 Chen) Anomalies == out of distribution features.

Weak localization or unsupervised segmentation via domain translation

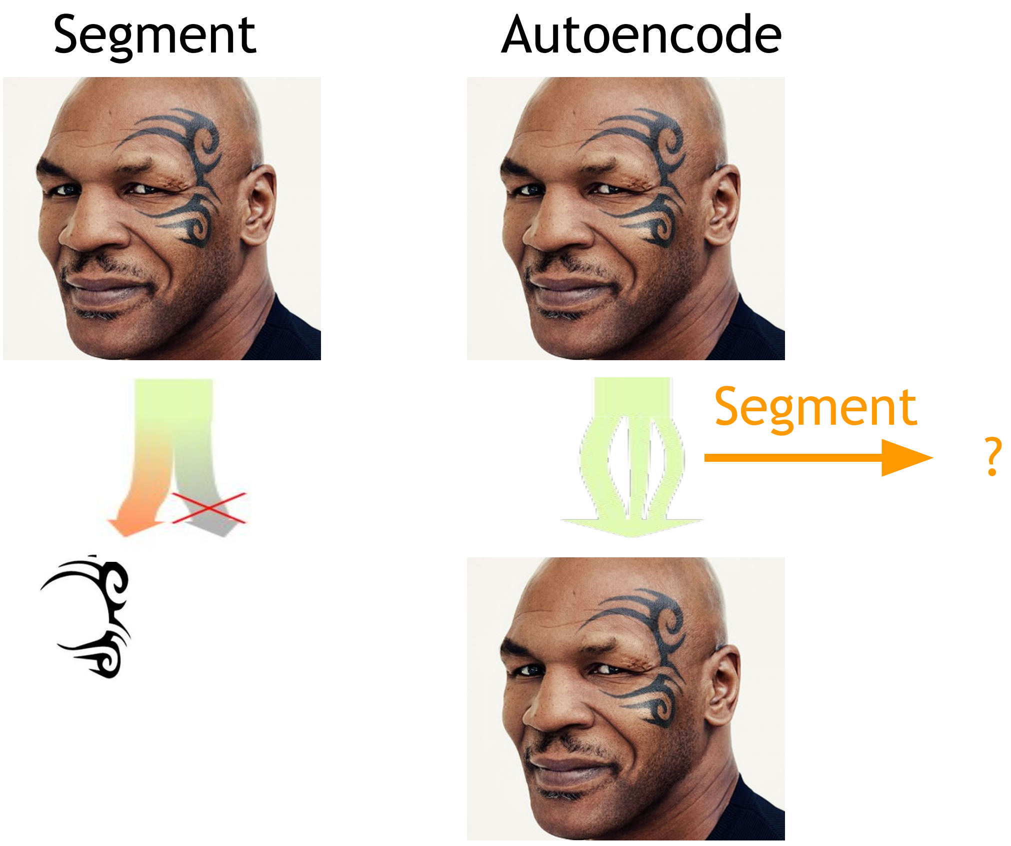
- (2017 Baumgartner) Sick → Healthy; find area of greatest change.
- (2018 Andermatt) Sick → Healthy → Sick; predict infilling region.

Zhang et al. "Deep adversarial networks for biomedical image segmentation utilizing unannotated images." MICCAI 2017 conference. Springer, Cham, 2017.
Schlegl et al. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery." IPMI 2017 conference. Springer, Cham, 2017.
Baur et al. "Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images." arXiv preprint arXiv:1804.04488 (2018).
Chen et al. "Unsupervised Detection of Lesions in Brain MRI using constrained adversarial auto-encoders." arXiv preprint arXiv:1806.04972 (2018).
Baumgartner et al. "Visual feature attribution using Wasserstein GANs." CVPR 2017 conference.
Andermatt et al. "Pathology Segmentation using Distributional Differences to Images of Healthy Origin." arXiv preprint arXiv:1805.10344 (2018).

Method

Method

Choosing the right objective



SUPERVISED

- Segmentation disentangles TATTOO from TYSON.

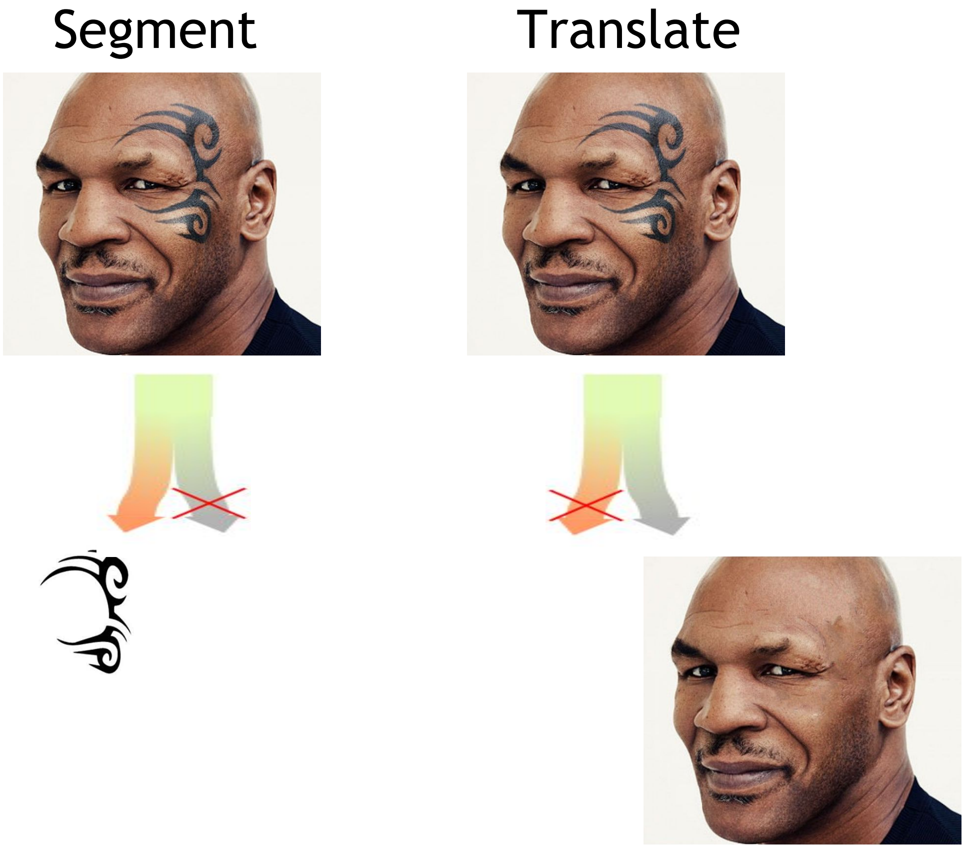
UNSUPERVISED

- Autoencoder disentangling may differ.

Which unsupervised loss disentangles like segmentation?

Method

Choosing the right objective



- Segment : keep TATTOO.
- Translate : keep TYSON.

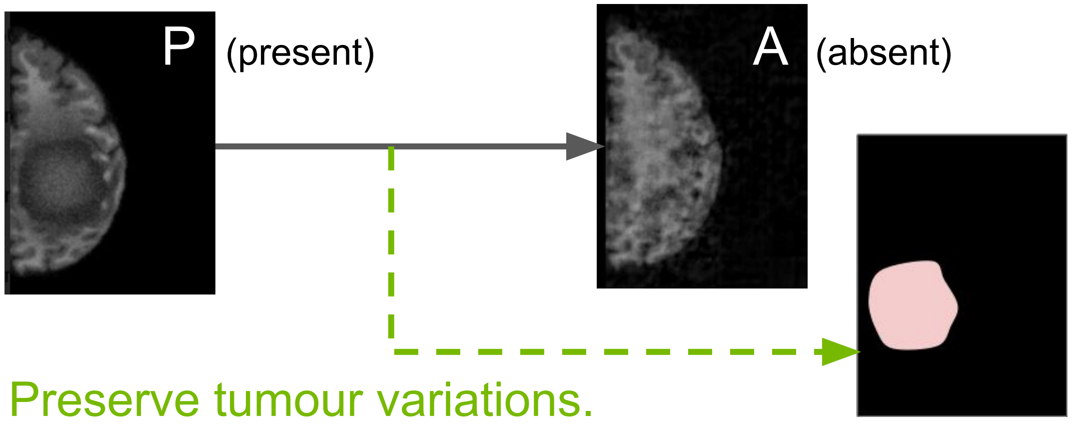
Method

Choosing the right objective

Combine: autoencode

- Translate but keep all variations.
- Recombine to autoencode.

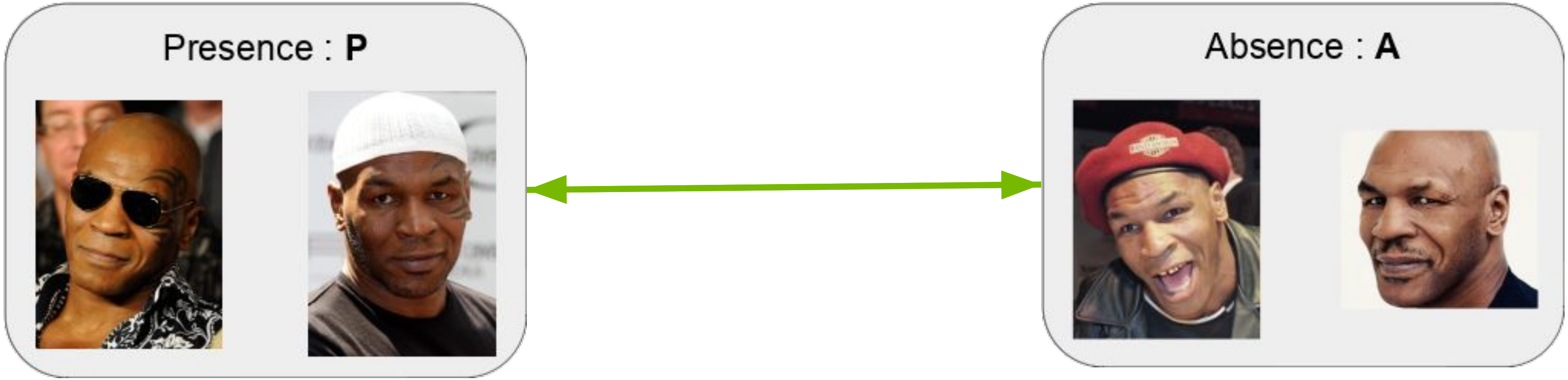
Example: sick to healthy translation



Method

Unpaired translation

As in (M)UNIT, (Augmented) CycleGAN
(2017 Liu, 2018 Huang, 2017 Zhu & Park, 2018 Almahairi)

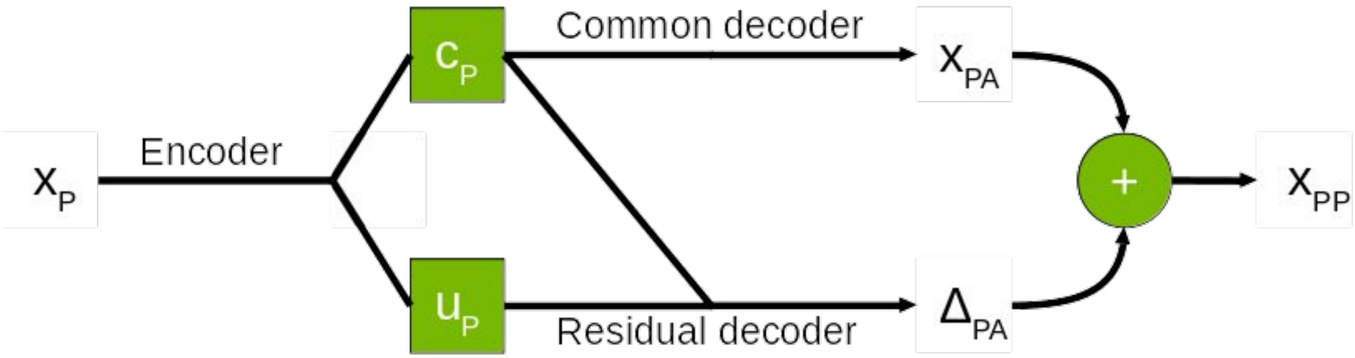


Liu et al. "Unsupervised image-to-image translation networks." NIPS 2017 conference.
Huang et al. "Multimodal Unsupervised Image-to-Image Translation." arXiv preprint arXiv:1804.04732 (2018).
Zhu & Park et al. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks." ICCV 2017.
Almahairi et al. "Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data." arXiv preprint arXiv:1802.10151 (2018).

P : presence
A : absence

Method

Translation model

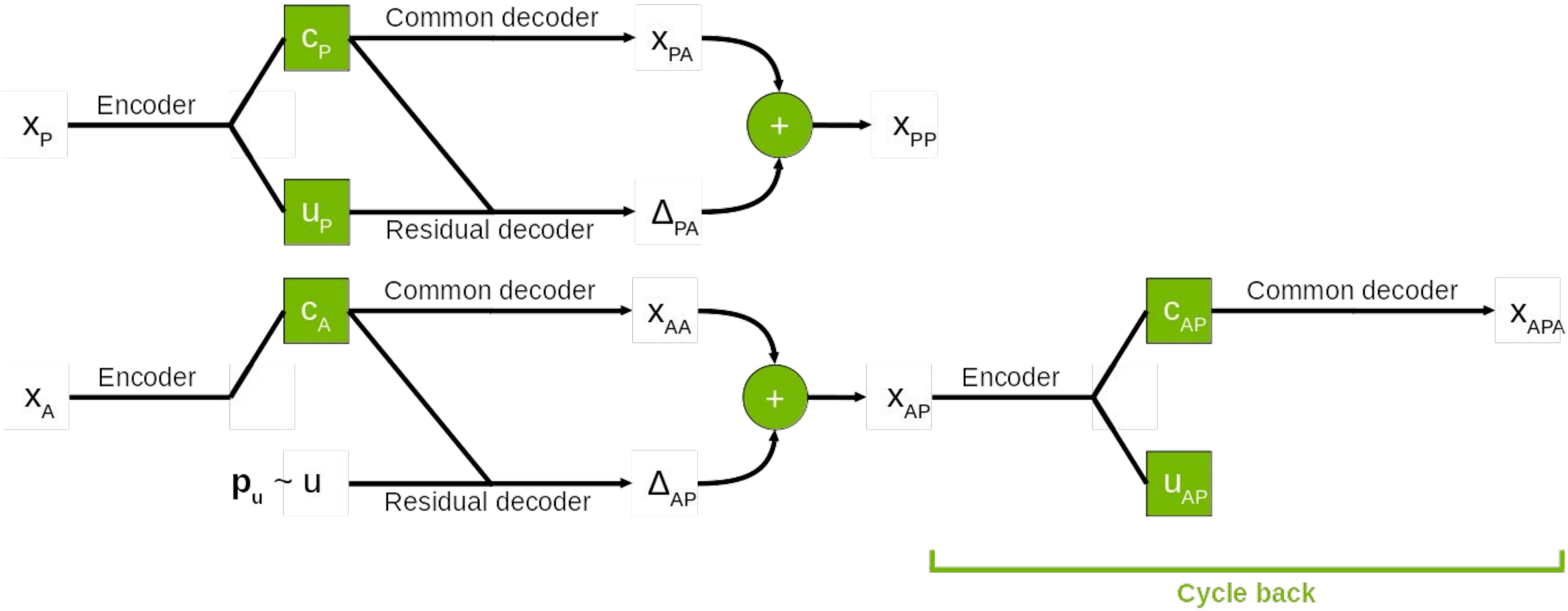


c : common features (in P and A)
u : unique features (in P)

P : presence
A : absence

Method

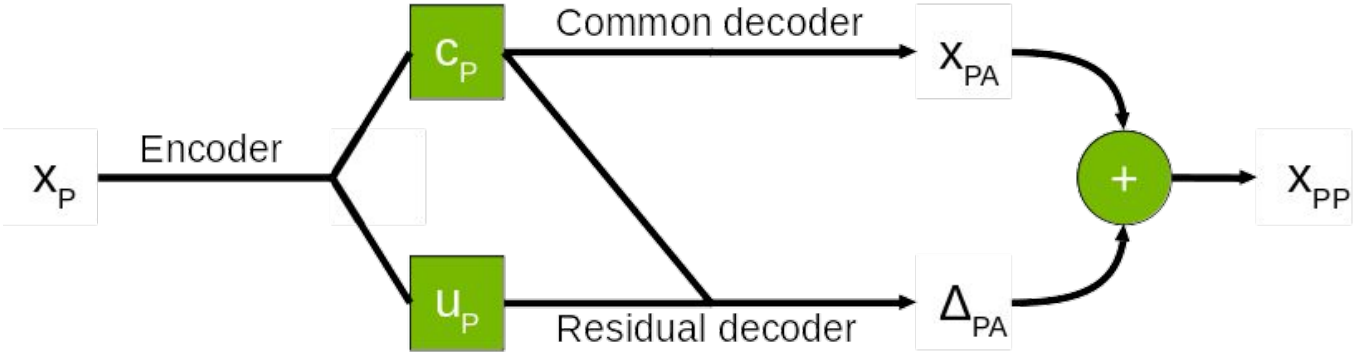
Translation model



P : presence
A : absence

Method

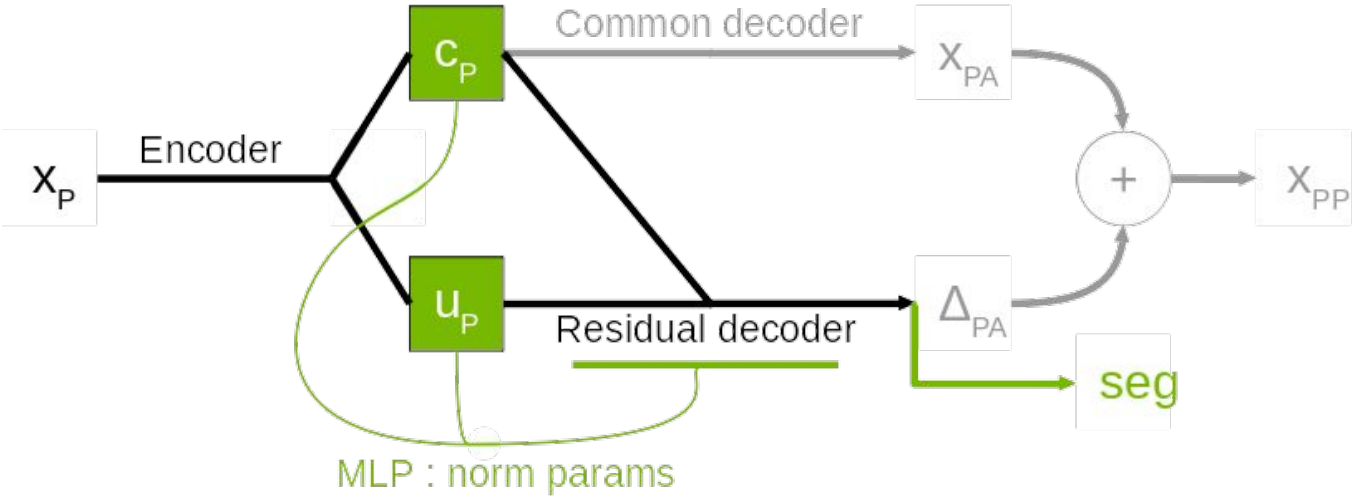
Segmentation : re-use decoder



P : presence
A : absence

Method

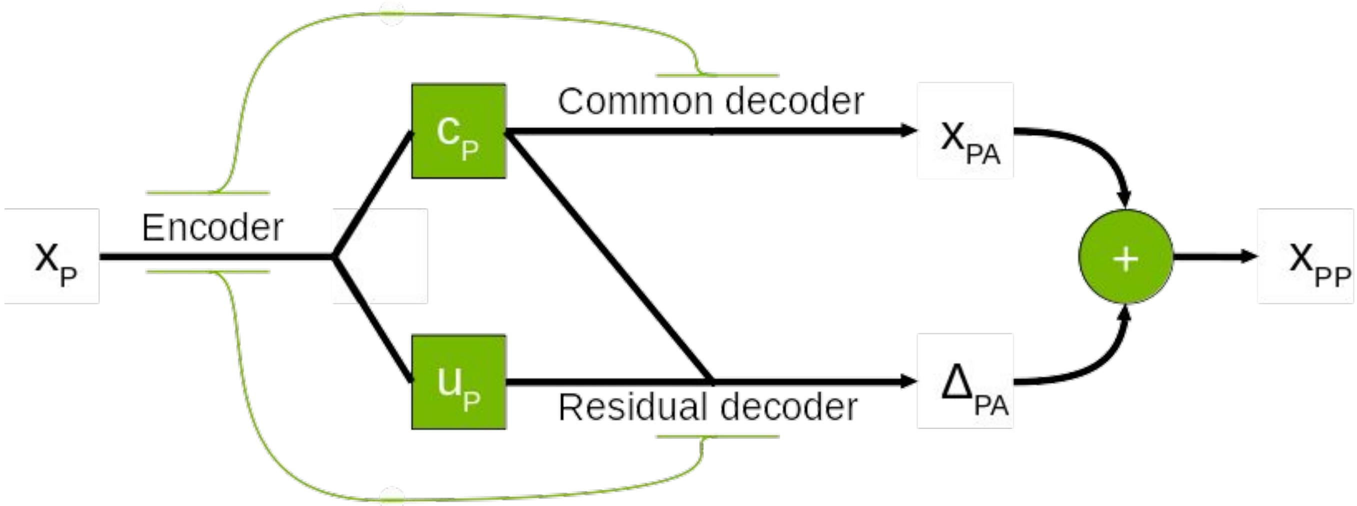
Segmentation : re-use decoder



P : presence
A : absence

Method

Recover spatial detail : long skip connections



Long skip: compress (1x1 conv) & concatenate

Method

Training objectives

Reconstruct image: $L_1(x_P, x_{PP}) + L_1(x_A, x_{AA}) + L_1(x_P, x_{PAP})$

Reconstruct latent: $L_1(c_P, c_{PA}) + L_1(c_A, c_{AP}) + L_1(c_A, c_{AA}) + L_1(c_P, c_{PP}) + L_1(u_P, u_{PP}) + L_1(u, u_{AP})$

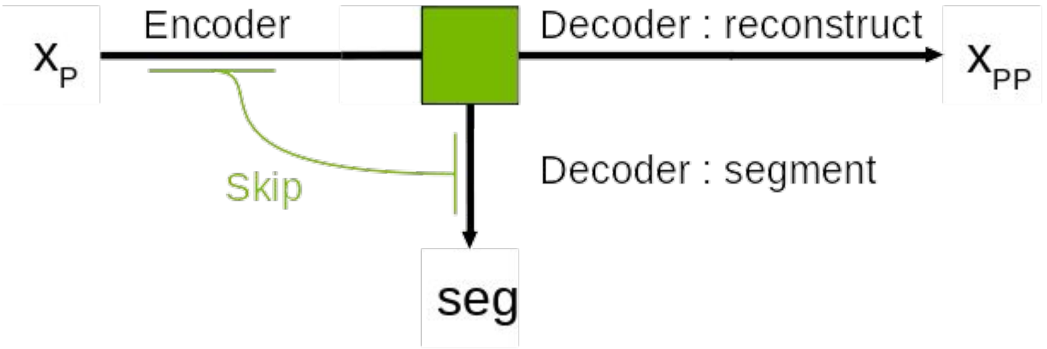
Adversarial: $-D_P(x_P, x_{AP}) - D_A(x_A, x_{PA})$

Segmentation: Dice(seg)

Evaluation

Evaluation

Semi-supervised baseline



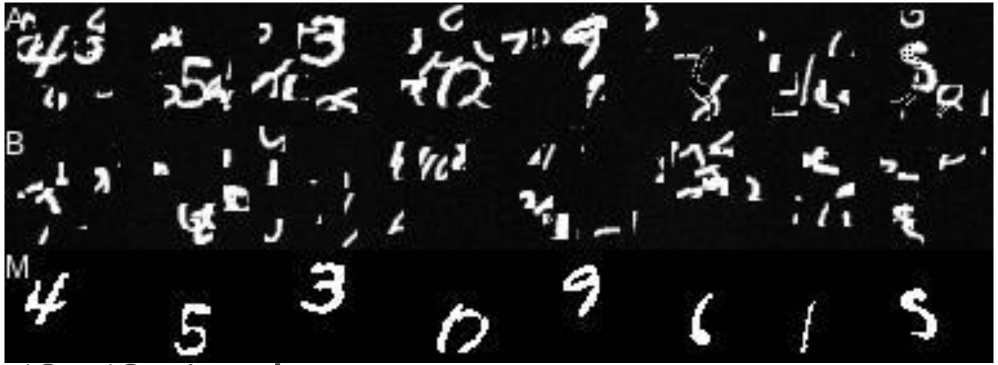
UNSUPERVISED: autoencode
SUPERVISED: segment

(Similar to the winning approach for BRATS 2018 by Andriy Myronenko, Nvidia)

P : presence
A : absence
M : segmentation mask

Evaluation

Synthetic data



48x48 simple

48x48 hard

128x128

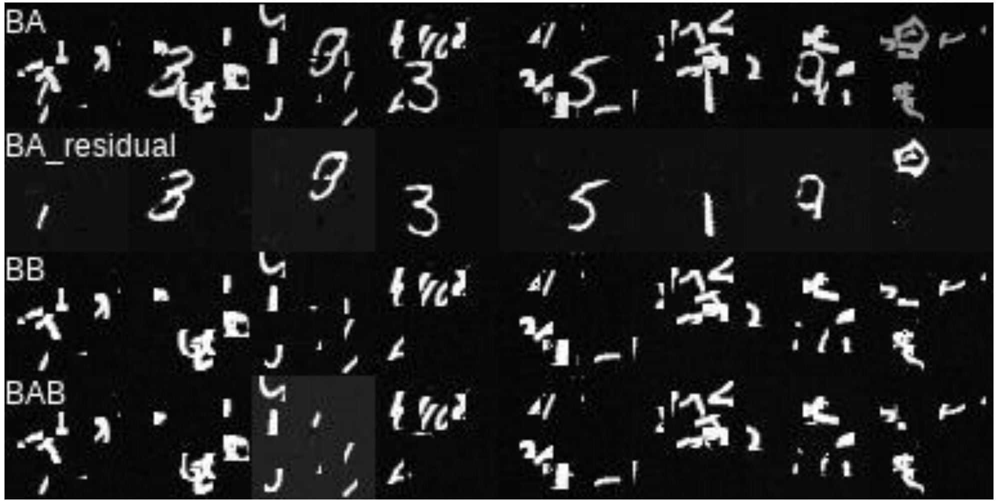
Using 1% of available segmentations.
Using segmentations only for digit 9.

P : presence
A : absence
M : segmentation mask

Evaluation

Synthetic data : 48x48 simple

M
PM
PA
Δ_{PA}
PP



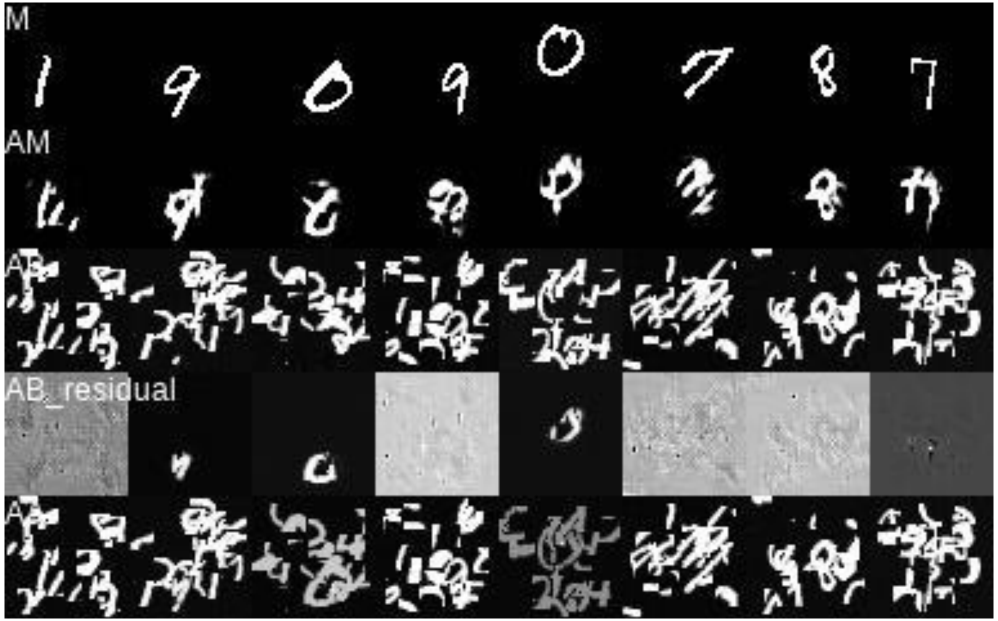
AP
Δ_{AP}
AA
APA

Using 1% of available segmentations.
Using segmentations only for digit 9.

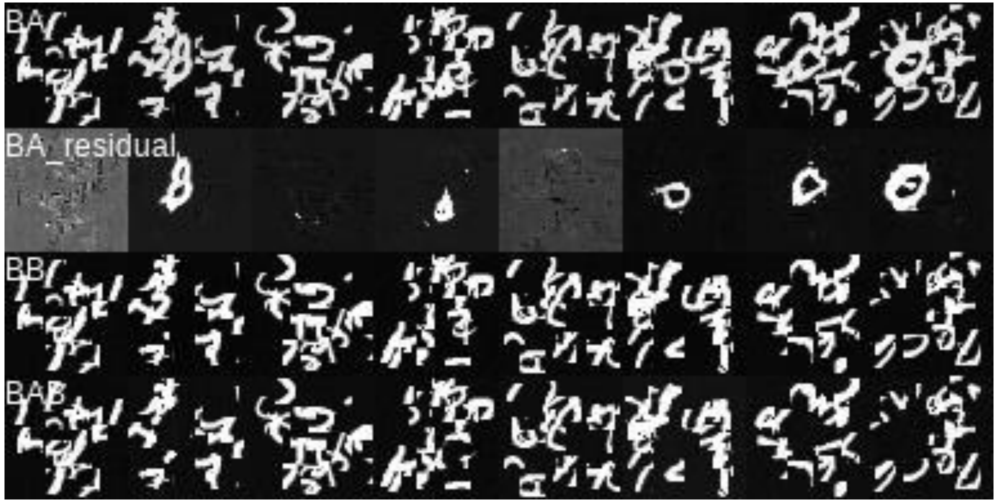
P : presence
A : absence
M : segmentation mask

Evaluation

Synthetic data : 48x48 hard



M
PM
PA
Δ_{PA}
PP



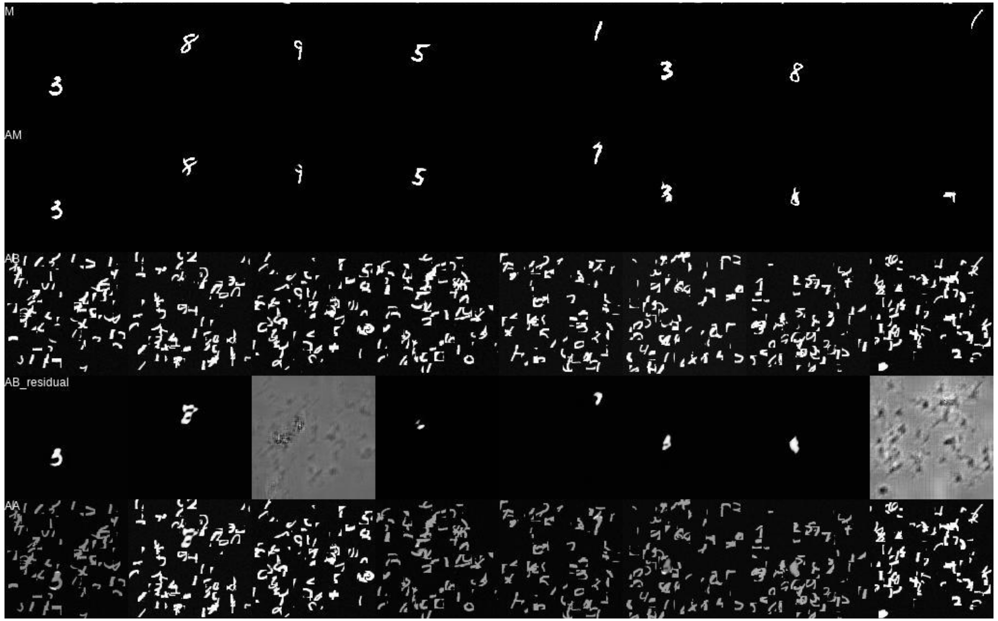
AP
Δ_{AP}
AA
APA

Using 1% of available segmentations.
Using segmentations only for digit 9.

P : presence
A : absence
M : segmentation mask

Evaluation

Synthetic data : 128x128



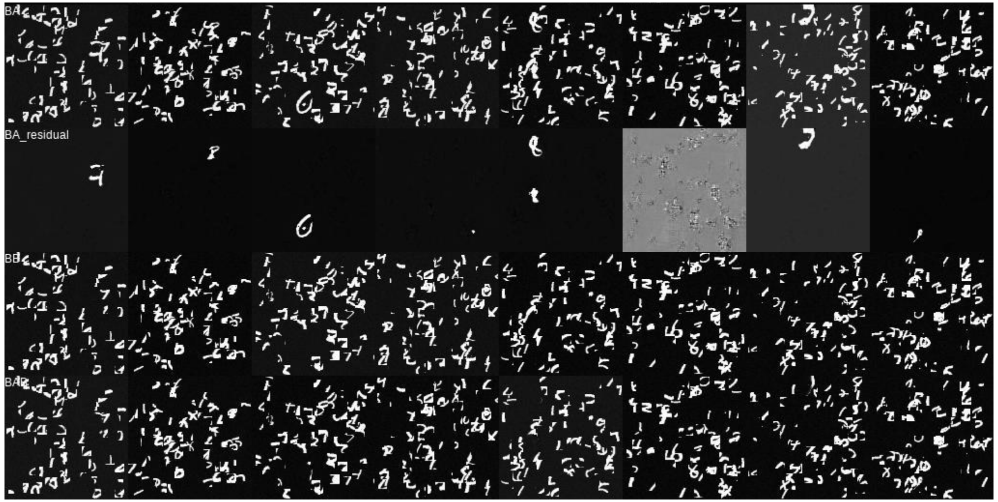
M

PM

PA

Δ_{PA}

PP



AP

Δ_{AP}

AA

APA

Using 1% of available segmentations.
Using segmentations only for digit 9.

Dice score as “mean (stdev)” - each over 3 runs.

Synthetic task (1% labels, digit 9) : DICE

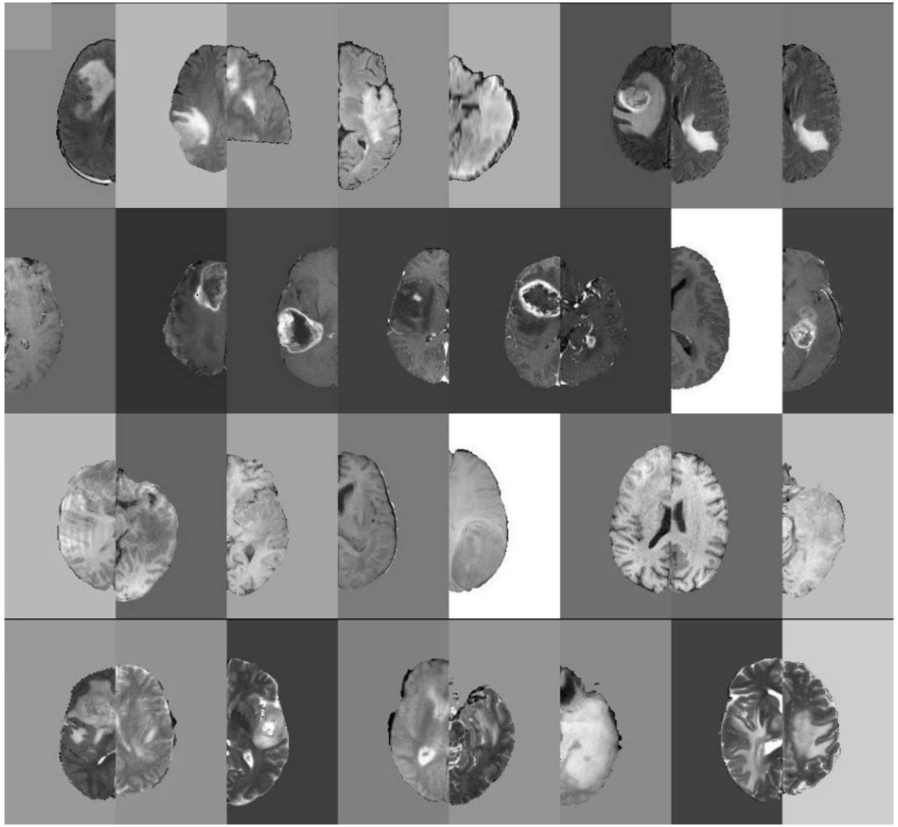
	48x48 simple	48x48 hard	128x128
Only segmentation	0.61 (0.01)	0.36 (0.01)	0.15 (0.01)
AE baseline	0.75 (0.01)	0.49 (0.02)	0.57 (0.02)
Proposed	0.79 (0.01)	0.57 (0.00)	0.65 (0.01)

P : presence
A : absence

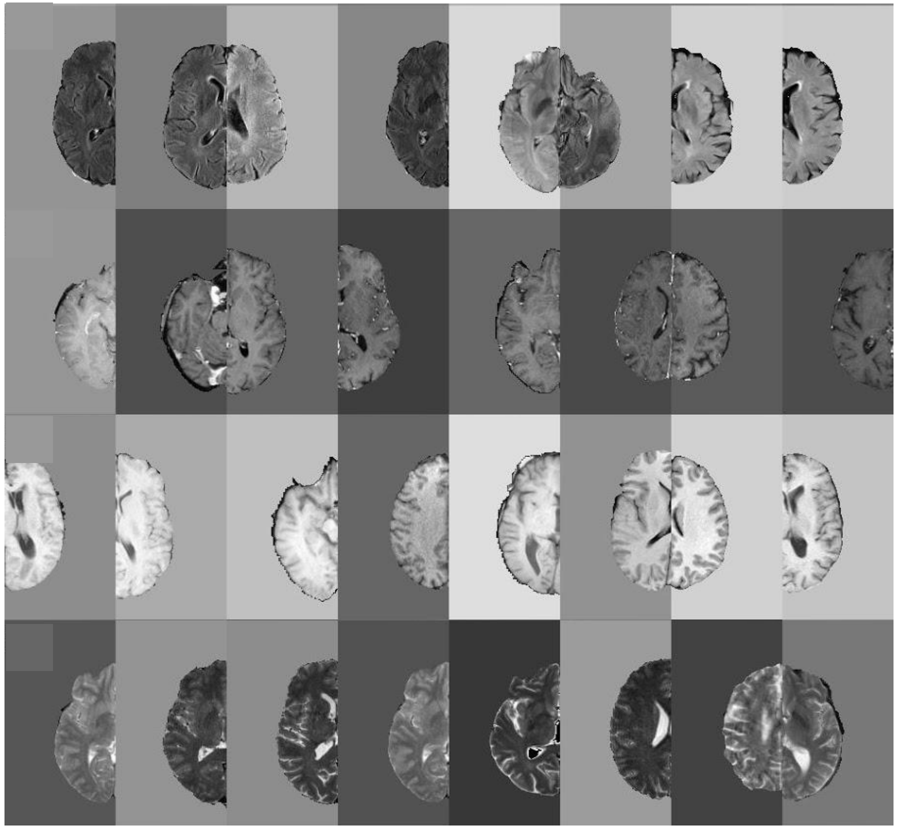
Evaluation

Brain tumor segmentation in MRI (BRATS 2017)

Flair
T1
T1c
T2



P



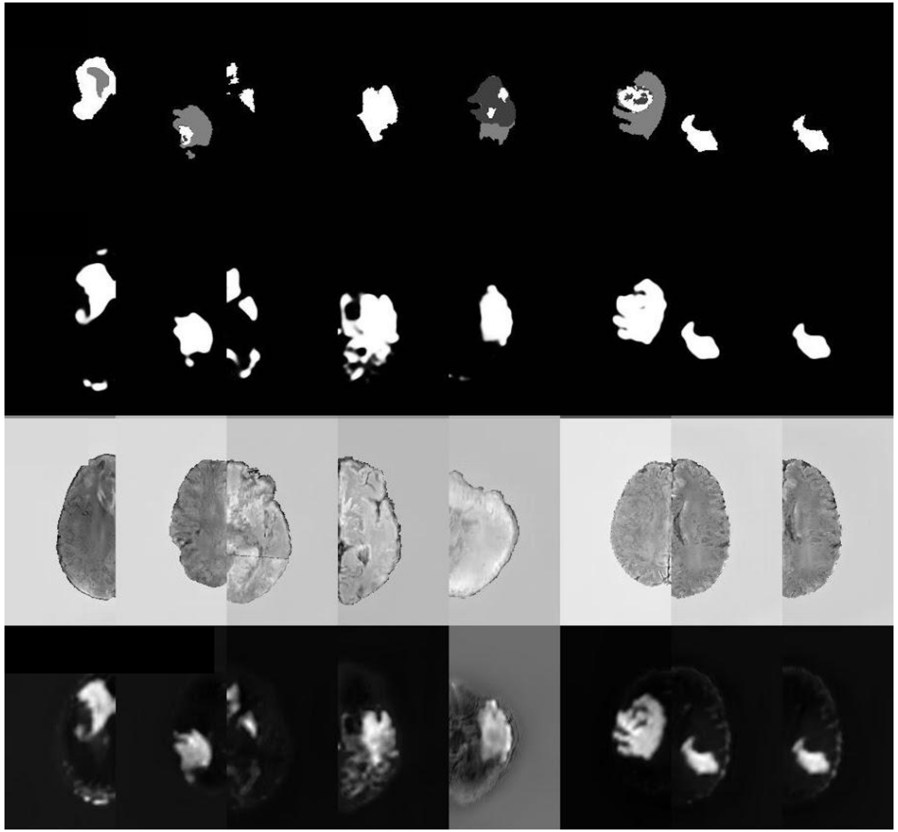
A

Using 1% of available segmentations.
P : presence
A : absence
M : segmentation mask

Evaluation

Brain tumor segmentation in MRI (BRATS 2017)

Flair

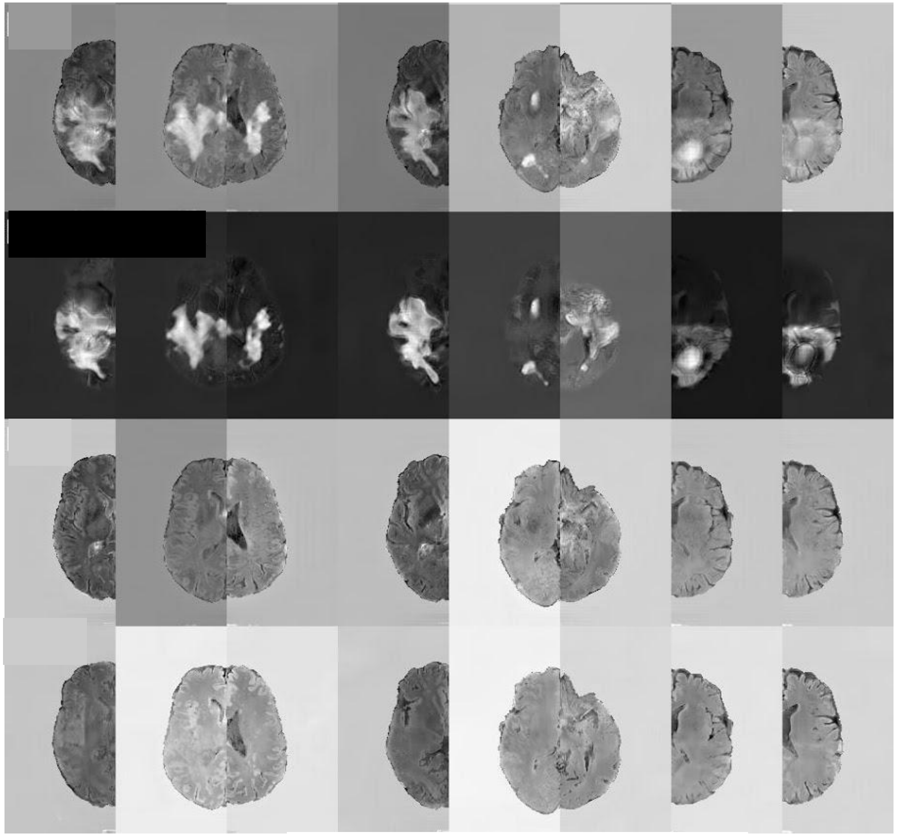


M

PM

PA

Δ_{PA}



AP

Δ_{AP}

AA

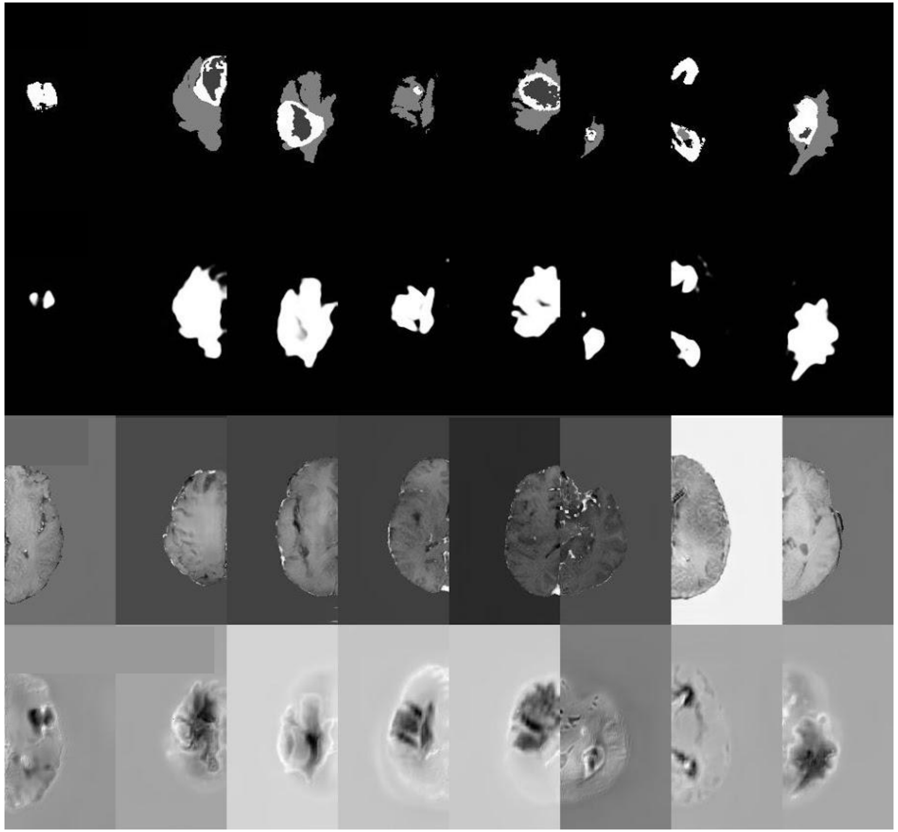
APA

Using 1% of available segmentations.
P : presence
A : absence
M : segmentation mask

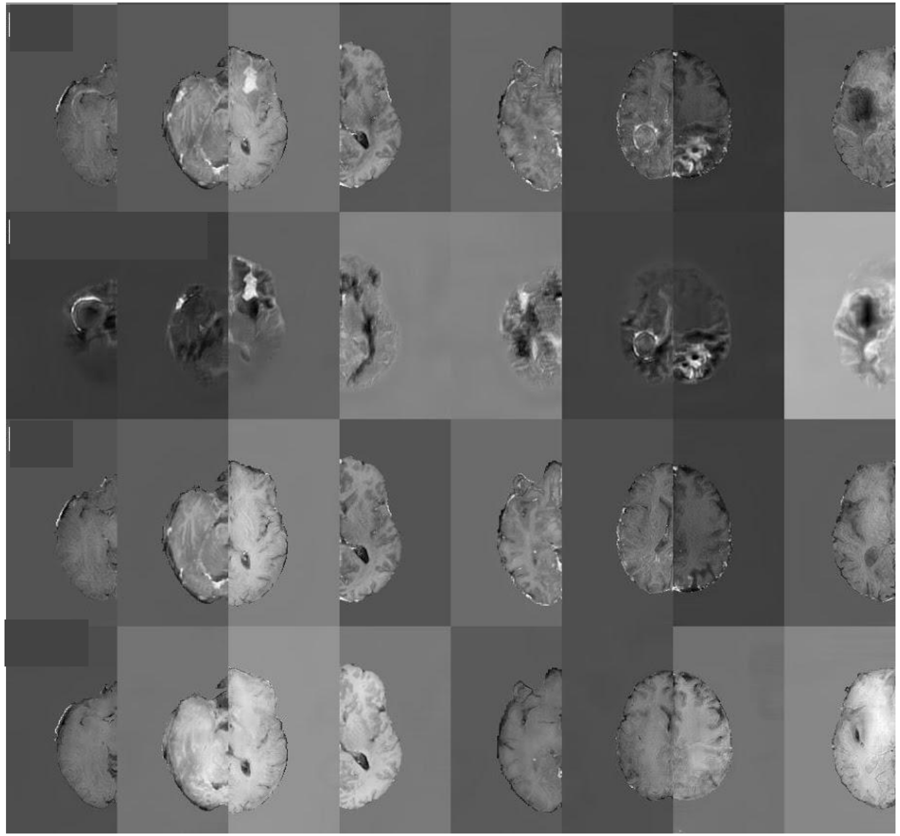
Evaluation

Brain tumor segmentation in MRI (BRATS 2017)

T1



M
PM
PA
Δ_{PA}

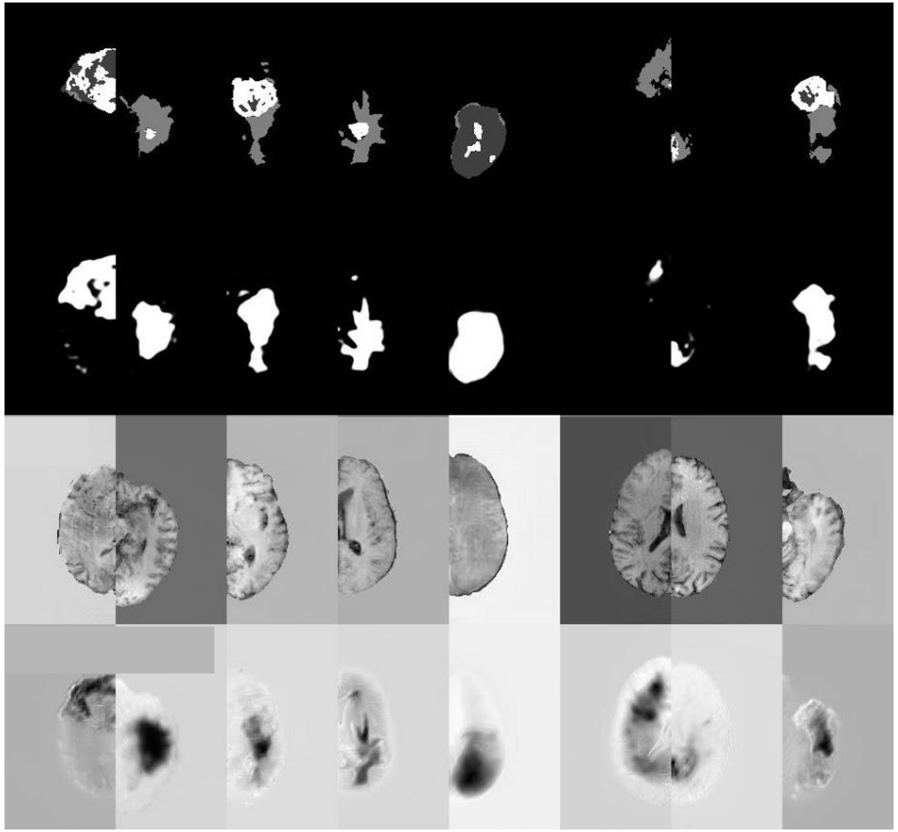


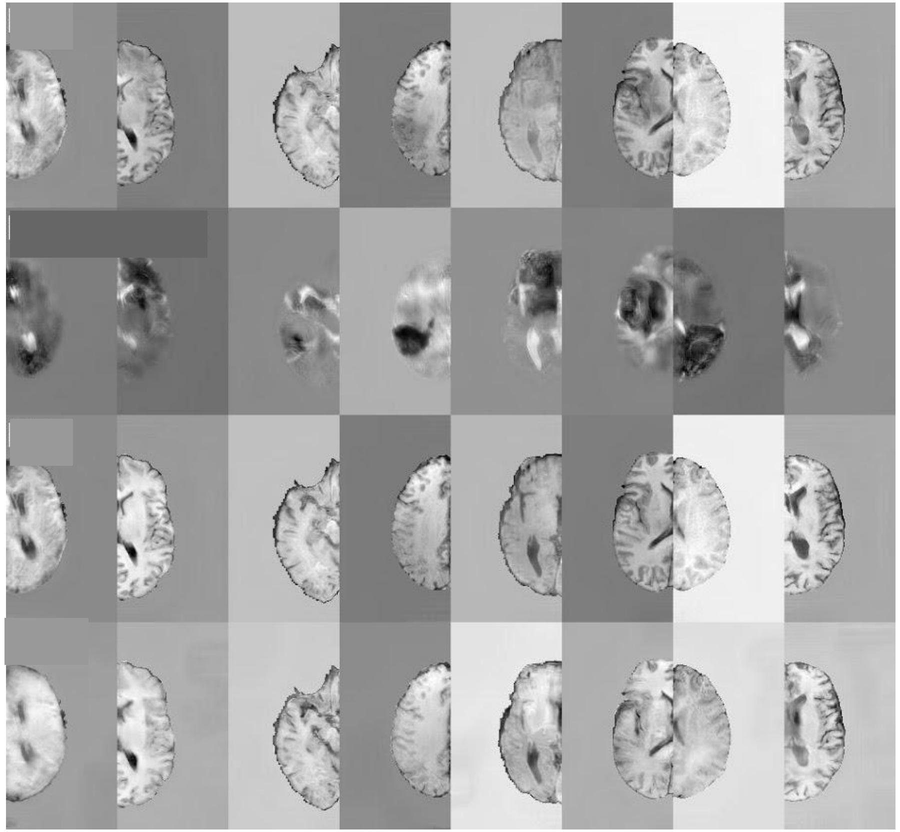
AP
Δ_{AP}
AA
APA

Using 1% of available segmentations.
P : presence
A : absence
M : segmentation mask

Evaluation

Brain tumor segmentation in MRI (BRATS 2017)

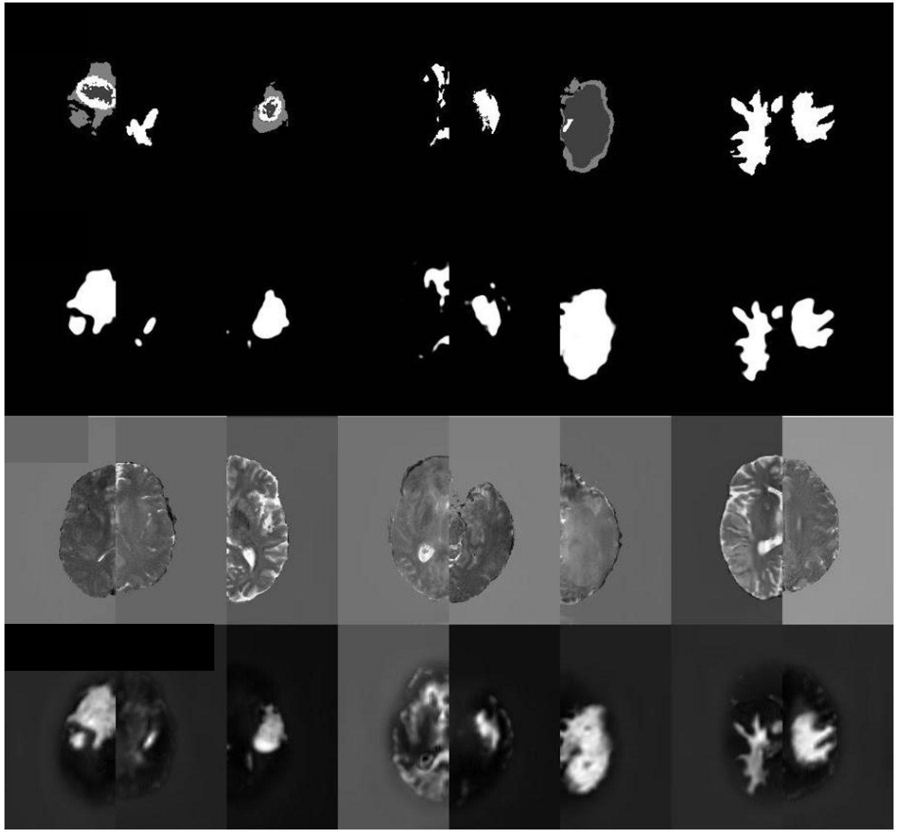




Using 1% of available segmentations.
P : presence
A : absence
M : segmentation mask

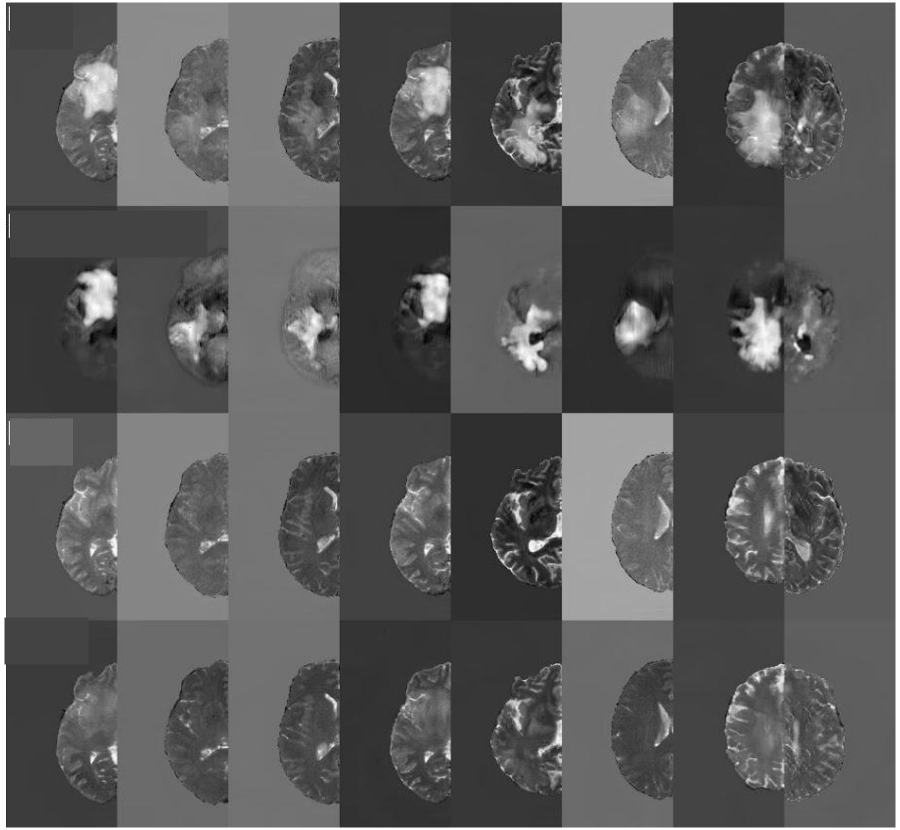
Evaluation

Brain tumor segmentation in MRI (BRATS 2017)



M
PM
PA
Δ_{PA}

T2



AP
Δ_{AP}
AA
APA

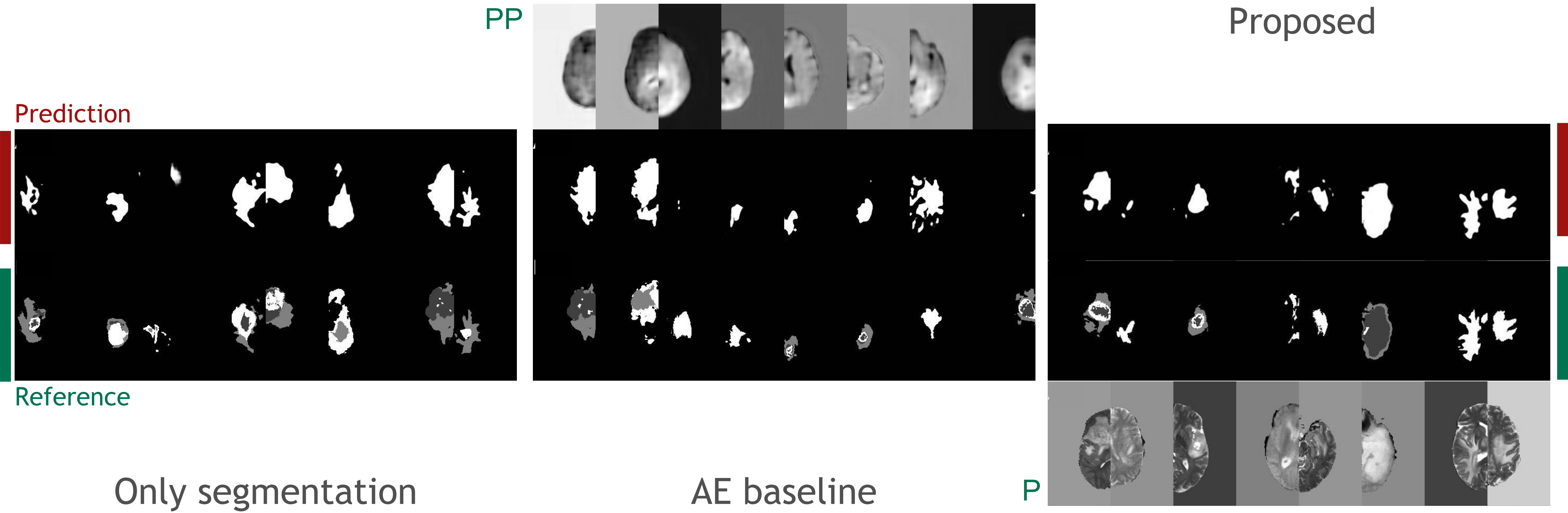
Dice score as “mean (stdev)” - each over 3 runs.

BRATS (1% labels) : DICE

	48x48 simple
Only segmentation	0.69 (0.04)
AE baseline	0.73 (0.02)
Proposed	0.79 (0.02)

Evaluation

BRATS segmentation comparison



Only segmentation

AE baseline

Disentangling

Disentangling

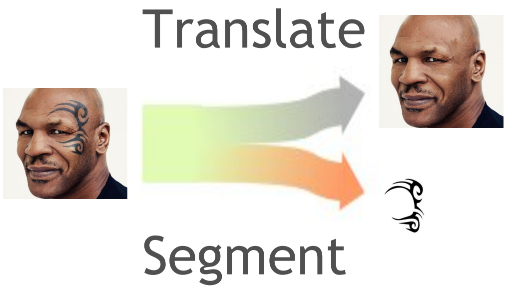
Separate domain-specific variations

Translation & segmentation disentangle same variations (common vs unique).

- (c) common: contains at least information in A
- (u) unique: contains at least information in P

Should **c** $\perp\!\!\!\perp$ **u** ?

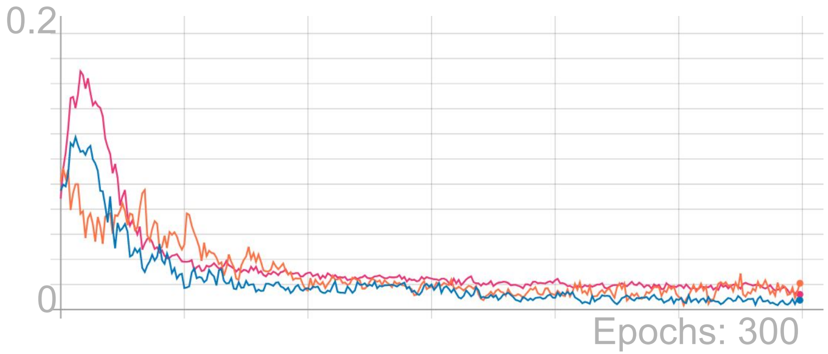
- If causes of unique features independent of causes of common features, in image.



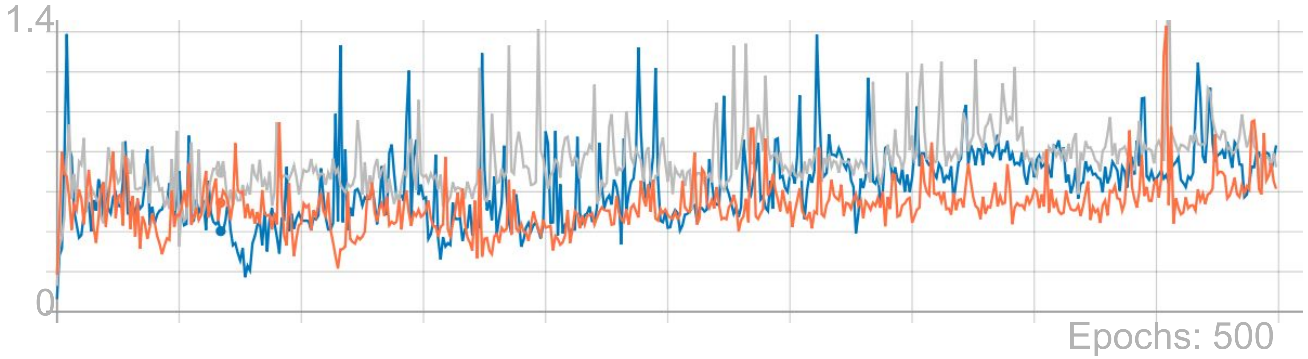
Disentangling

Mutual information measured

$\mathbf{c} \perp\!\!\!\perp \mathbf{u} : \min \mathrm{MI}(\mathbf{c}, \mathbf{u})$



Synthetic



BRATS

Disentangling

Mutual information minimized

$\mathbf{c} \perp\!\!\!\perp \mathbf{u}$: min MI($\mathbf{c}$, $\mathbf{u}$)

Work in process …

- Using mutual information neural estimator (MINE)
- May try gradient reversal layer, as in (2018 Gonzalez-Garcia)

Belghazi et al. "MINE: mutual information neural estimation." arXiv preprint arXiv:1801.04062 (2018).
Gonzalez-Garcia et al. "Image-to-image translation for cross-domain disentanglement." arXiv preprint arXiv:1805.09730 (2018).
Ganin et al. "Unsupervised Domain Adaptation by Backpropagation." ICML 2015 conference.

Further work

Further work

Further validation and robustness

Try more data

- Additional real data (DDSM, diabetic retinopathy)

Explore

- Enforce independence between **common** and **unique** (helpful?)

Technical

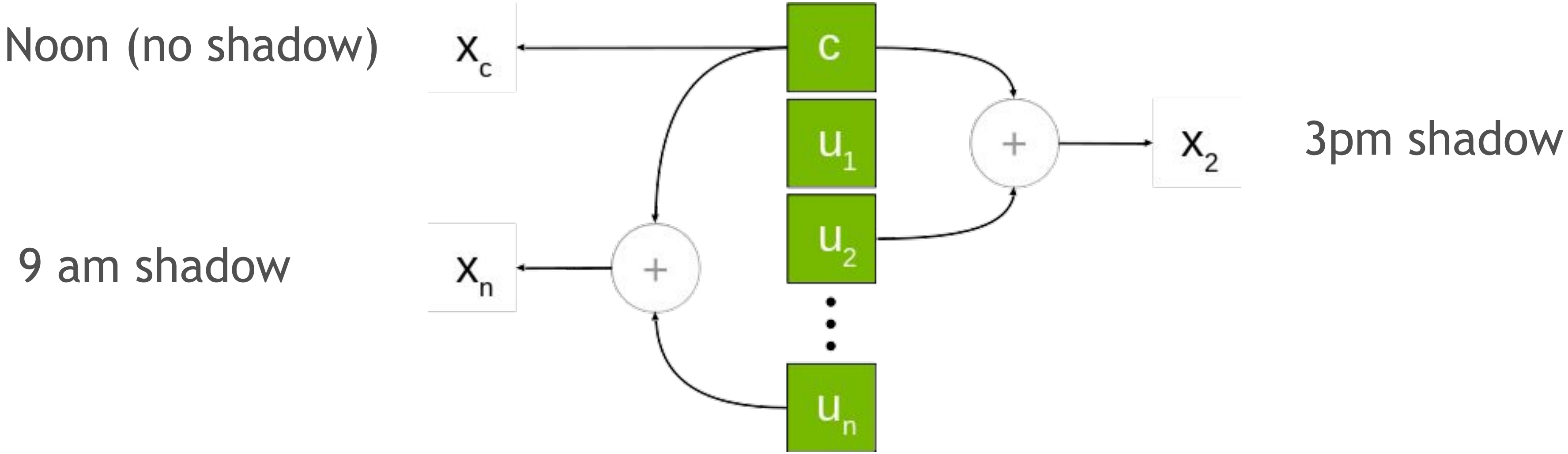
- Separate segmentation decoder (currently OK for MNIST, bad for BRATS)
  → Try pretraining non-segmentation objectives.
- Long skip : max pooling indices (pretrain autoencoding).

Extend

- Multi-domain translation

Further work

Multi-domain translation : example



Example 1: A scene at different times of day. Segment the shadows.

Example 2: Chest x-rays with different pathologies. Also healthy x-rays.

NVIDIA