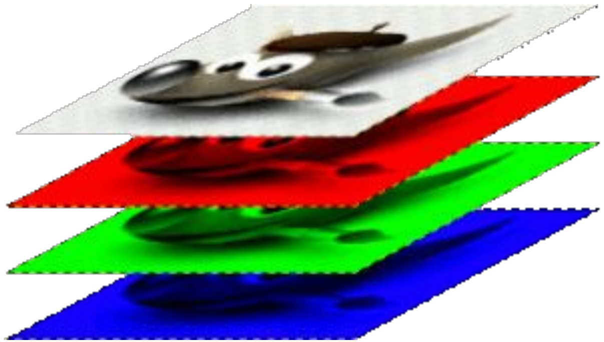
Dealing with Missing Modalities in Medical Images: What's Missing?

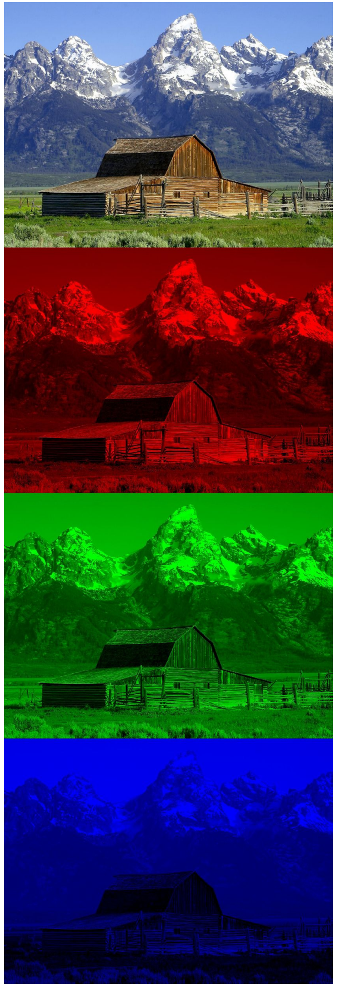
Anmol Sharma

Problem of Missing Modalities

Computer Vision Perspective

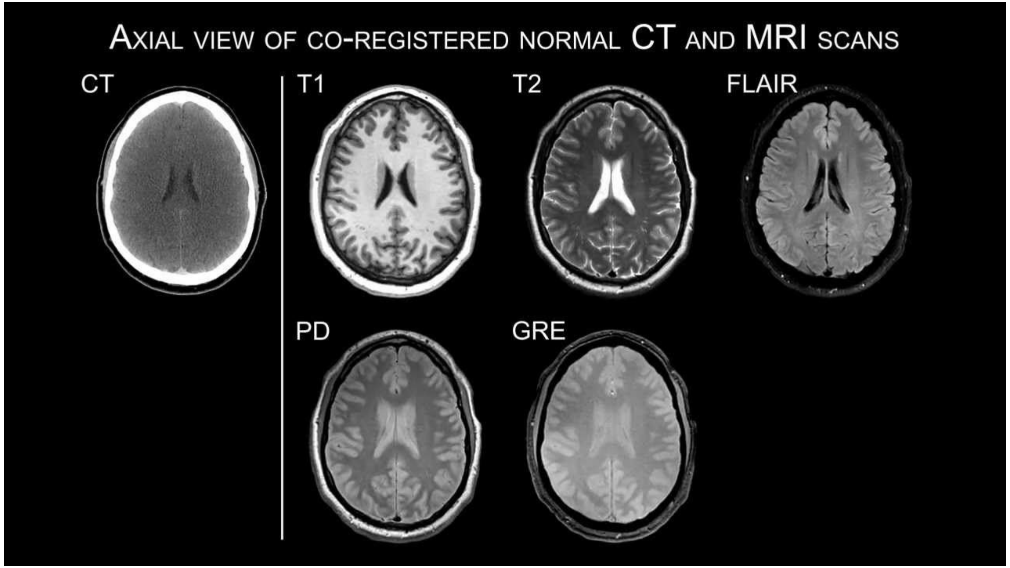
- Every image is a combination of 3 channels: RGB
- Most deep learning networks expect all three channels to be present to perform their task:
  - Classification
  - Detection
  - Instance Segmentation
- All 3 channels are typically acquired by every camera, so there is no significant issue.





Problem of Missing Modalities
Medical Imaging Perspective

- Some medical image types (MR) are acquired in multiple modalities or sequences^.
- For example: Brain MRI scans typically have 2-7 modalities acquired for each patient.
- The scans are acquired by adjusting relaxation times, and contrast values, so that some structures of interest are visible better in some modality.
- Modalities provide both redundant information, as well as complimentary information, which is leveraged by the physicians.

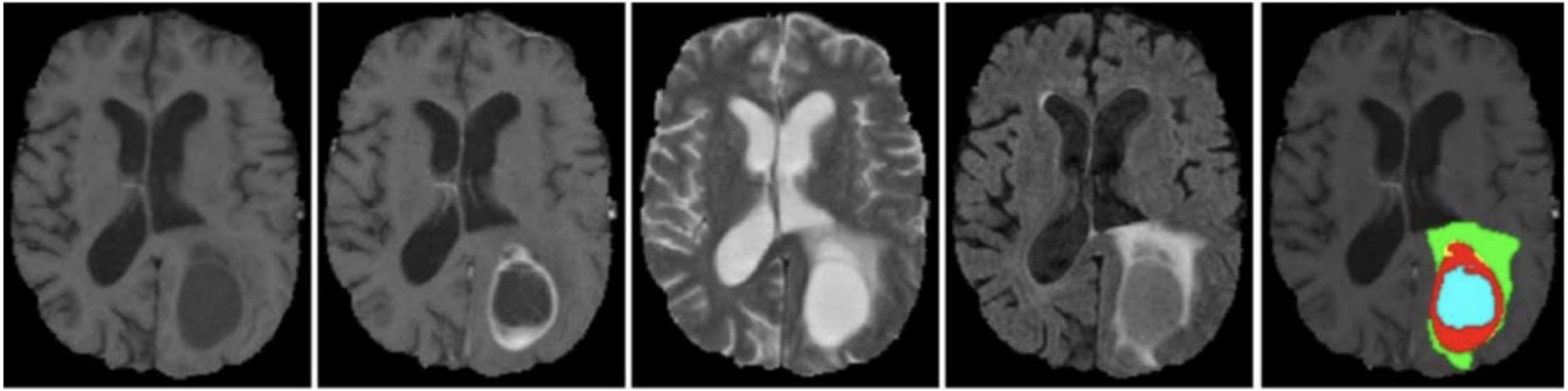


^Modality/Sequence are used interchangeably

Problem of Missing Modalities

Medical Imaging Perspective

Brain MRI Scans acquired in 4 modalities: T1, T1CE, T2FLAIR, T2
Each modality is an independent 3D volume



Usual Shape of this Data: (4, 240, 240, 155) (4D Data)
Where 4 = Number of modalities

So What’s the problem?
Availability of Clean Data and Hard Coded Dependence

- Availability of clean data has lead to strong dependence on it in most of the methods designed.
- BRATS dataset is a very clean and organized dataset with a set of 4 modalities (T1, T2, T1CE, T2FLAIR).
- All the approaches designed for brain tumor segmentation/classification that were built upon BRATS data strongly require the above modalities to work.
- This makes it hard to generalize the approaches (or translate a learnt model) to other data, which may or may not contain all the above modalities.

So What's the problem?

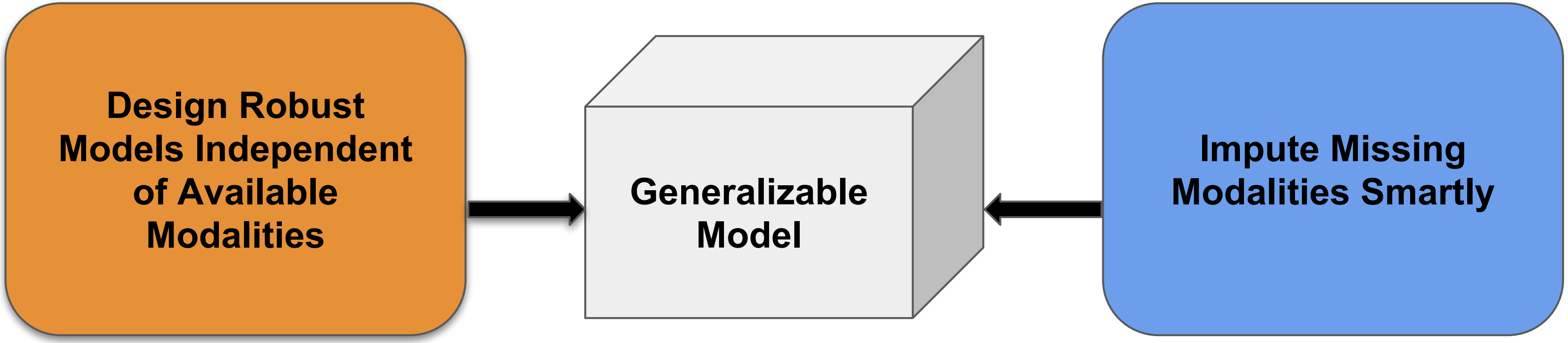
Every Hospital has their own Image Acquisition Protocols

- Translating learnt models/atlases which were designed on BRATS data towards real world hospital data is very tricky, due to their dependence on a fixed set of 4 modalities.
- An incorrect modality type, or a missing one usually throws the method off and renders it almost useless.
- **Every hospital has a different imaging protocol**
  - Some collect T1/T2 regularly, and only collect T2FLAIR if required.
  - Some may collect new sequences like T1Gd, MPRAGE, and even Diffusion Weighted (DWI).

<u>What do we do when we don't have a modality that our model requires to work?</u>

<u>How do we generalize our models to work with any data, regardless of the missing modalities?</u>

Approaches to the Problem



Approaches to Problem

Design Robust Models

- Heteromodal Image Segmentation (HeMIS)
  - Havaei, Mohammad, et al. "HeMIS: Hetero-modal image segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2016.

- Permutation Invariant Multimodal Segmentation (PIMMS)
  - Varsavsky, Thomas, et al. "PIMMS: Permutation Invariant Multi-Modal Segmentation." (2018).

Design Robust Models Independent of Available Modalities

Approaches to Problem

Impute/Generate Missing Modality

- All previous literature dealing with imputing missing data.
    - Classic methods like:
        - Mean/Median/Mode Imputation
        - Mutual Information Based Imputation
        - Other custom methods

- Multimodal MR Synthesis via Modality-Invariant Latent Representation
    - Chartsias, Agisilaos, et al. "Multimodal mr synthesis via modality-invariant latent representation." *IEEE transactions on medical imaging* 37.3 (2018): 803-814.

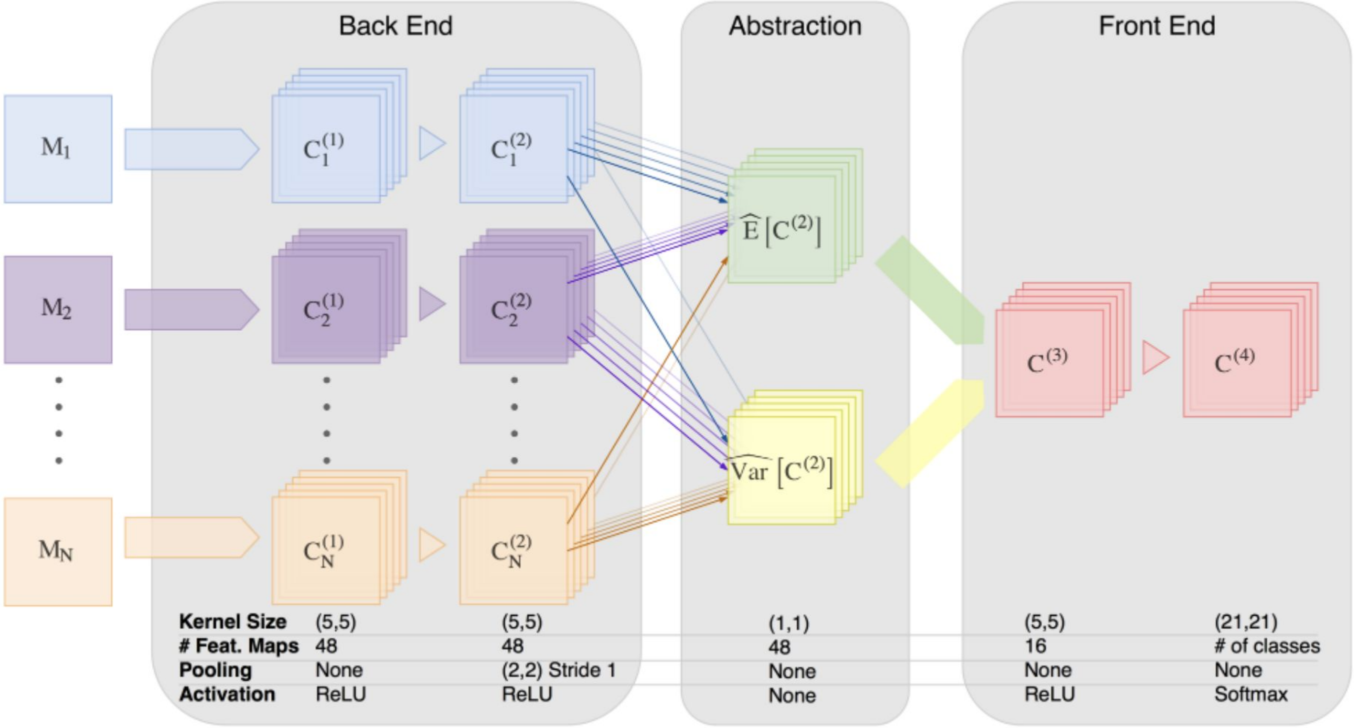
Impute Missing Modalities Smartly

Design Robust Models

HeMIS: Hetero-modal image segmentation

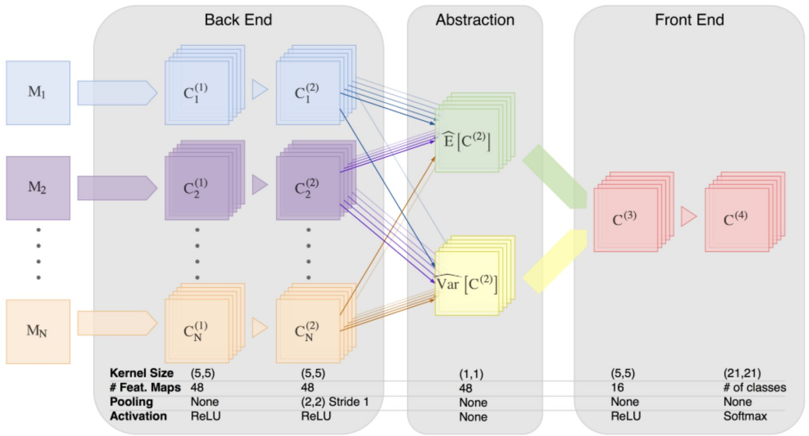
Mohammad Havaei^, Nicolas Guizard^, Nicolas Chapados^, Yoshua Bengio*
^Imagia Inc. Montreal Canada
*Université de Montréal Montréal Canada

HeMIS Model Architecture



HeMIS
Model Architecture

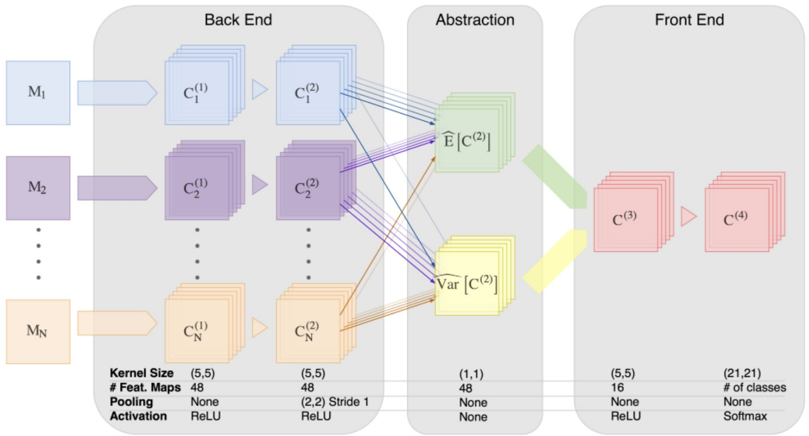
- Back End
  - Initial “encoders” working independently for each modality.
  - Each encoder is built using 2 convolutional layers.
  - A final pooling layer with Stride = 1 is used, which preserves shape.
- Abstraction
  - Combine information from the independent encoders into a shared representation.
  - The abstraction operation must be invariant to the cardinality of input data.
    - Mean and variance is defined for any *n* numbers.
    - This provides robustness to the network as it doesn’t depend on a lowest common multiple of modalities to work.



HeMIS

Model Architecture

- Front End
  - Decoder architecture with two convolutional layers.
  - Gets input as the shared representation generated in the abstraction layer (mean and variance feature maps).
  - Final softmax layer to generate masks corresponding to each class.



HeMIS

Training Model

- Pseudo Curriculum Learning
  - For initial “warm-up” epochs, show the network all modalities.
  - Eventually, start dropping one or more modalities (with higher probability of dropping one).
- Task
  - Segmentation of structure of interest (tumor, MS lesion)
- Loss Function
  - Pixel wise cross entropy loss function
- Data Used
  - MSGC Dataset
    - 20 MR Cases
    - T1W, T2W, T2FLAIR
  - RRMS MR Dataset
    - 300 Patients
    - T1W, T2W, T1C
  - BRATS Dataset
    - 220 Patients
    - T1, T2, T1CE, T2FLAIR

Results
BRATS 2013

Table 1. Comparison of HeMIS when trained on all modalities against BRATS-2013 Leaderboard and Challenge winners, in terms of Dice Similarity (scores from [13]).

	Leaderboard			**Challenge**		
Method	Complete	Core	Enhancing	Complete	Core	Enhancing
Tustison [21]	79	65	53	87	**78**	**74**
Zhao [25]	79	59	47	84	70	65
Meier [13]	72	60	53	82	73	69
HeMIS	**83**	**67**	**57**	**88**	75	**74**

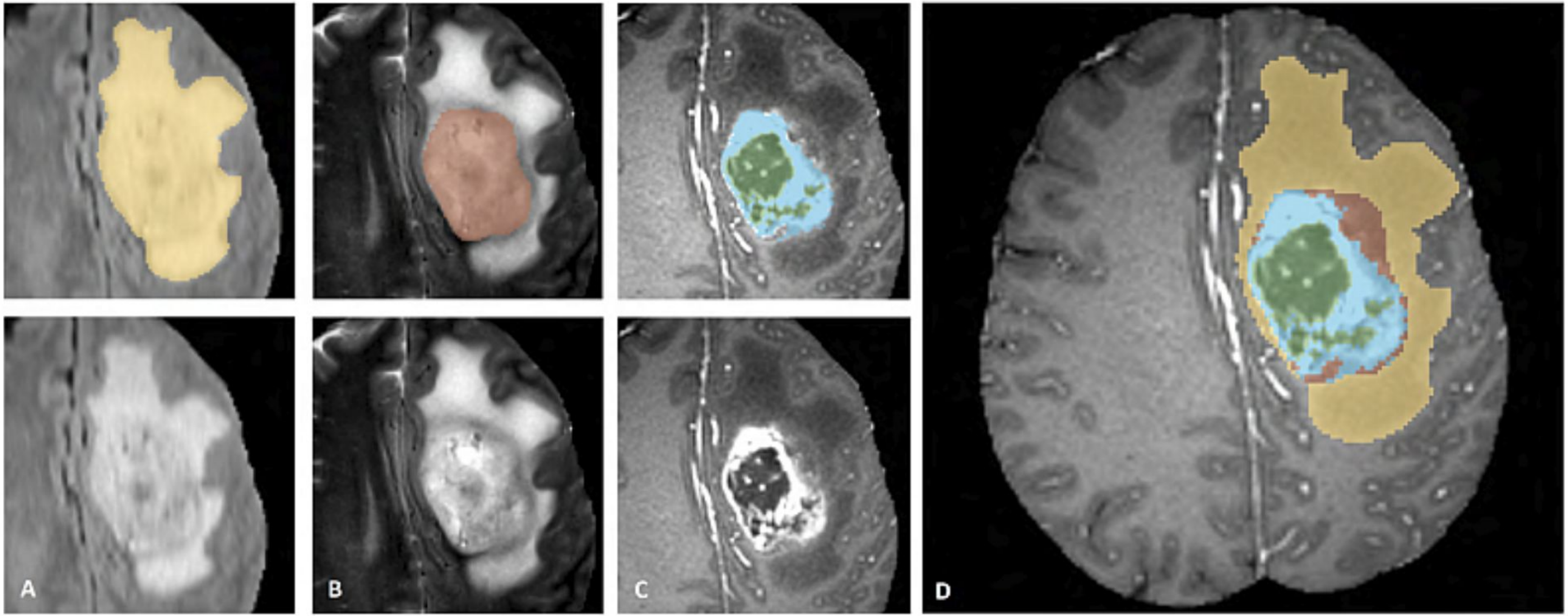
Results

MSGC Dataset

Table 2. Results of the full dataset training on the MSGC. For each rater (CHB and UNC), we provide the volume difference (VD), surface distance (SD), true positive rate (TPR), false positive rate (FPR) and the method's score as in [18].

Method	Rater	VD (%)	SD (mm)	TPR (%)	FPR (%)	Score
Souplet et al. [16]	CHB	86.4	8.4	58.2	70.6	80.0
	UNC	57.9	7.5	49.1	76.3	
Geremia et al. [5]	CHB	**52.4**	**5.4**	59.0	71.5	82.1
	UNC	**45.0**	**5.7**	51.2	76.7	
Brosch et al. [2]	CHB	63.5	7.4	47.1	**52.7**	**84.0**
	UNC	52.0	6.4	**56.0**	**49.8**	
HeMIS	CHB	127.4	7.5	**66.1**	55.3	83.2
	UNC	68.2	6.6	52.3	61.3	

Results
BRATS and RRMS



Results
BRATS and RRMS

				RRMS			BRATS								
Modalities				Lesion			Complete			Core			Enhancing		
F	T_1	T_1c	T_2	HeMIS	Mean	MLP	HeMIS	Mean	MLP	HeMIS	Mean	MLP	HeMIS	Mean	MLP
○	○	○	●	1.74	2.66	**12.77**	58.48	2.70	**61.50**	**40.18**	4.00	37.32	**20.31**	6.25	18.62
○	○	●	○	2.67	0.00	**3.51**	**33.46**	23.11	2.04	**44.55**	23.90	17.70	**49.93**	30.02	32.92
○	●	○	○	3.89	0.00	**6.64**	**33.22**	0.00	2.07	**17.42**	0.00	10.52	4.67	6.25	**10.78**
●	○	○	○	34.48	9.77	**38.46**	71.26	**72.30**	63.81	**37.45**	0.00	34.26	5.57	6.25	**15.90**
○	○	●	●	**27.52**	4.31	25.83	**67.59**	35.01	64.97	**63.39**	30.92	49.38	**65.38**	39.00	60.30
○	●	●	○	8.21	0.00	**8.26**	**45.93**	23.63	1.99	**55.06**	41.89	26.55	**62.40**	43.80	40.93
●	●	○	○	38.81	11.62	**39.15**	**80.28**	75.58	78.13	**49.52**	0.00	48.97	22.26	6.25	**25.18**
○	●	○	●	**31.25**	8.31	29.39	**69.56**	1.77	66.88	**47.26**	2.63	43.66	23.56	6.25	**26.37**
●	○	○	●	**39.64**	33.31	38.55	**82.1**	81.01	81.35	**53.42**	25.94	52.41	23.19	6.25	**25.01**
●	○	●	○	**41.38**	6.42	39.33	79.8	45.97	**81.13**	**66.12**	29.85	65.51	**67.12**	35.14	66.19
●	●	●	○	**41.97**	9.00	40.63	80.88	81.57	**82.19**	69.26	62.13	**69.34**	**71.30**	67.13	70.93
●	●	○	●	**46.6**	41.12	41.83	**83.87**	77.84	80.40	**57.76**	20.66	53.46	**28.46**	6.25	28.34
●	○	●	●	**41.90**	38.95	41.47	82.78	64.19	**83.37**	**70.62**	42.36	70.45	70.52	49.62	**70.56**
○	●	●	●	**34.98**	5.78	29.46	**70.98**	30.86	67.85	**66.60**	45.79	55.40	**67.84**	50.21	64.81
●	●	●	●	**48.66**	43.48	43.48	**83.15**	82.43	82.43	**72.5**	71.46	71.46	**75.37**	72.08	72.08
# Wins / 15				**9**	0	6	**10**	1	4	**14**	0	1	**9**	0	6

Synthesize Missing Modalities

Multimodal MR Synthesis via Modality-Invariant Latent Representation

Agisilaos Chartsias, Thomas Joyce, Mario Valerio Giuffrida, and Sotirios A. Tsaftaris
University of Edinburgh, UK

Multimodal Synthesis
Background

- Each modality (T1, T2, T1CE, T2FLAIR) has redundant information.
- These modalities can be thought about as different functions with slight addition of an external factor.
- The hypothesis is that using a number of given modalities, a missing modality can be generated from the information gained from the available ones.

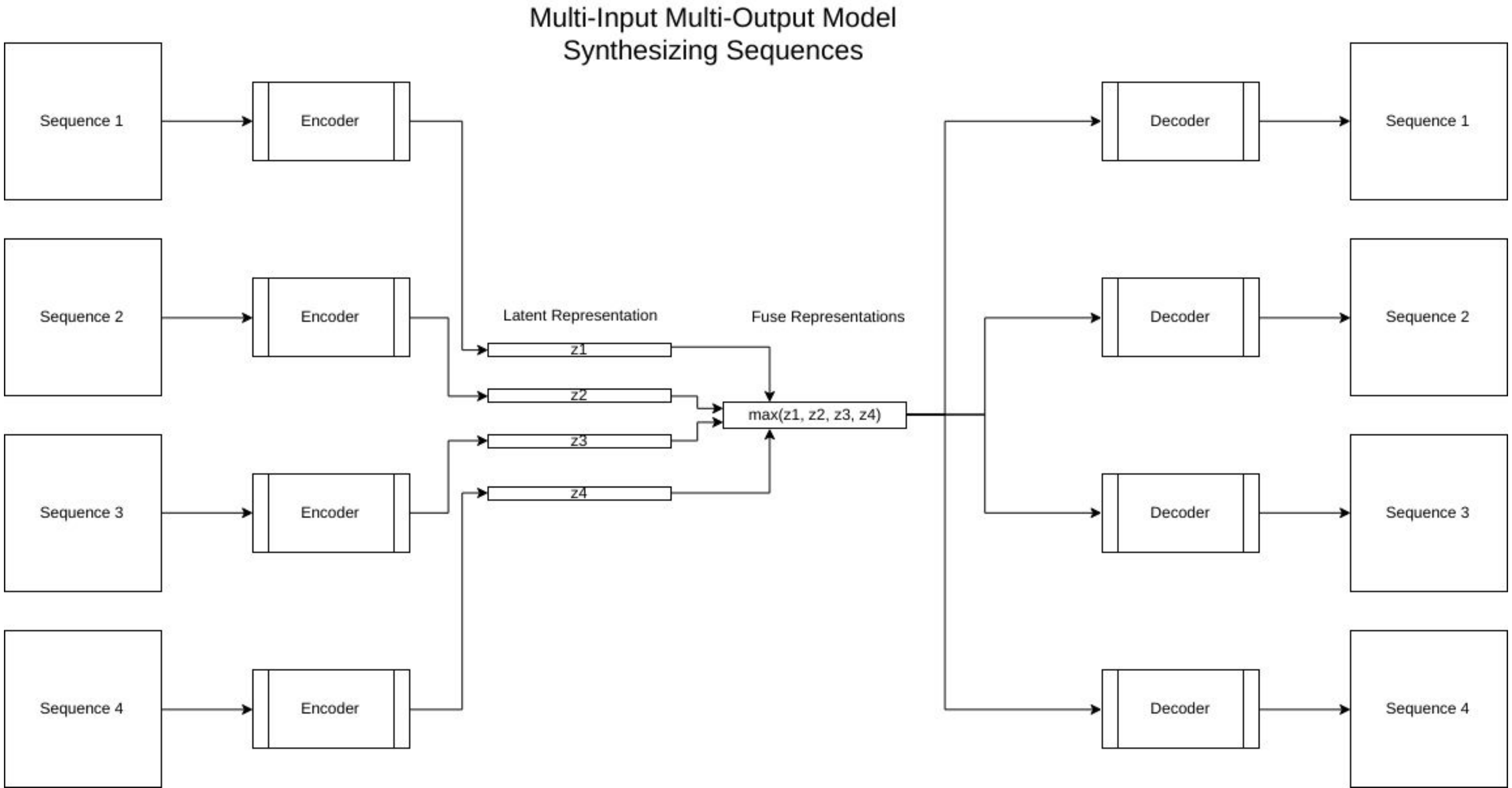
Multimodal Synthesis
Background

- However, to generate each sequence from any combination of available sequences, a total of $n(n-1)$ models would have to trained.
    - T1 -> T2
    - T2 -> T1
    - T1, T2 -> T2FLAIR
    - And so on…
- The naive approach doesn’t look too good in terms of computational load.

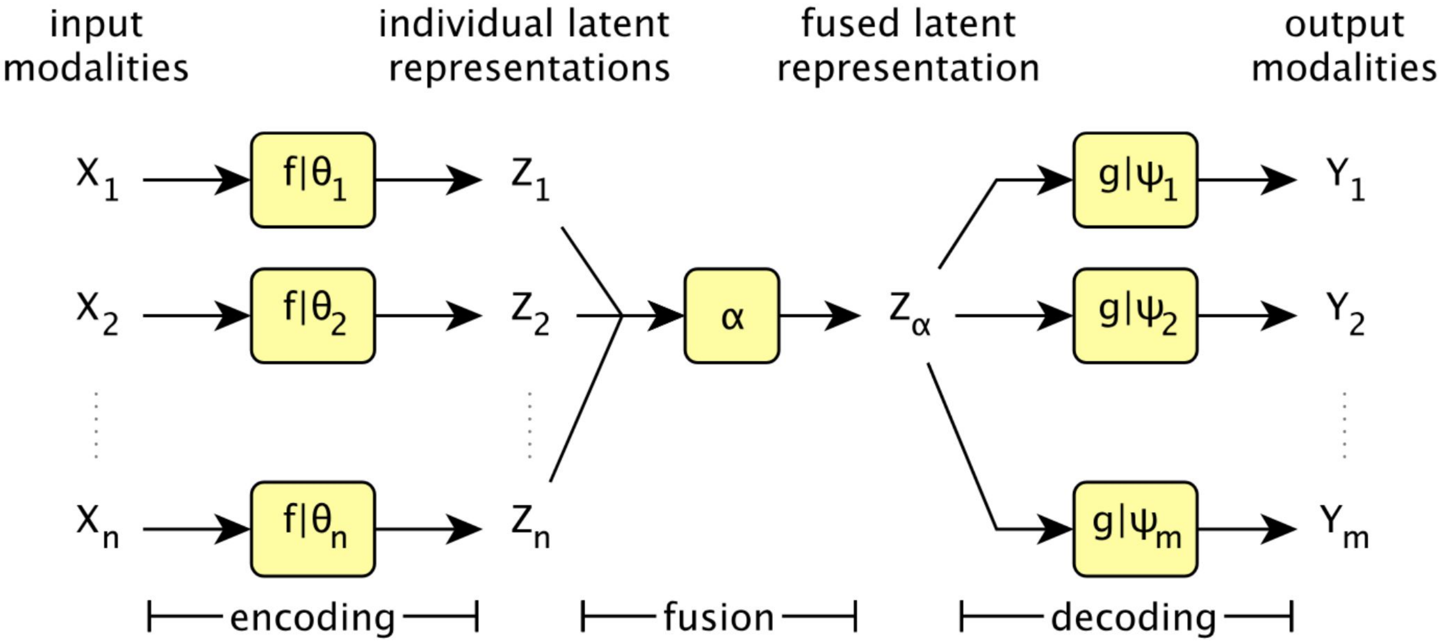
Multimodal Synthesis
Hypothesis

Design a model architecture that can scale to any number of given modalities, and generate any required modality.

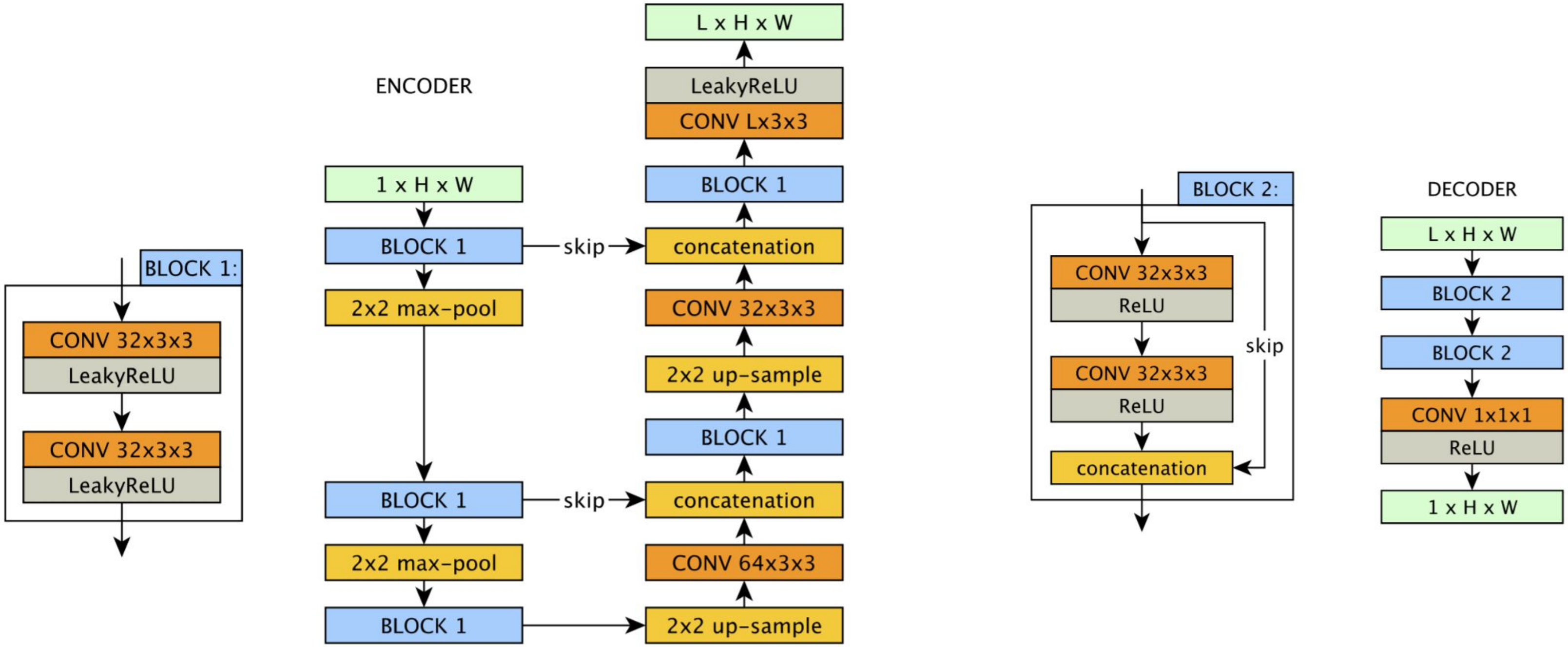
Multimodal Synthesis
Model Architecture



Multimodal Synthesis
Model Architecture

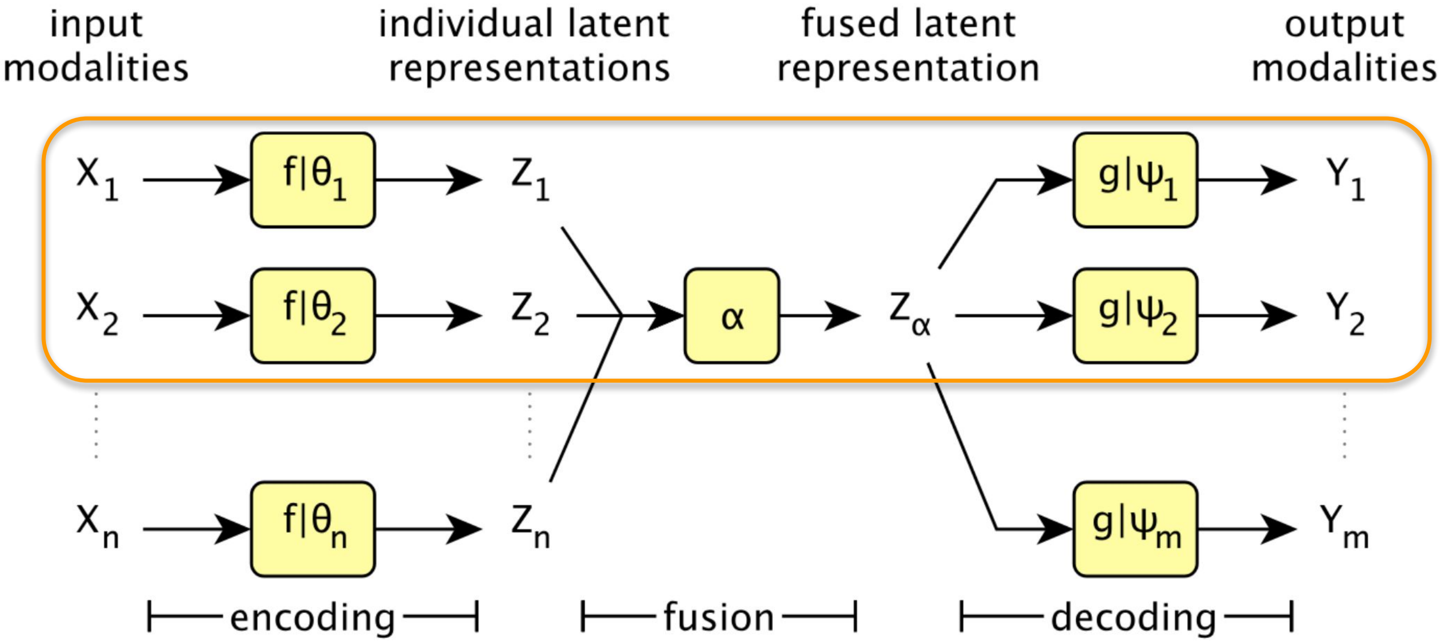


Multimodal Synthesis
Model Architecture



Multimodal Synthesis

Test Time Partial Model



Multimodal Synthesis
Training

- Trained end-to-end:
  - Specify the total available modalities as inputs
  - Specify the total available modalities as outputs
  - If we have 4 modalities, we train a model with 4 inputs and 4 outputs.
- No curriculum learning is used (random dropping of sequences).
- Data Used
  - BRATS 2015
    - 54 Patients from LGG
    - T1, T2, T1CE, T2FLAIR
    - Used only T1, T2 and T2FLAIR
  - ISLES 2015
    - 28 Patients
    - T1W, T2W, T2FLAIR, DWI
  - IXI Dateset
    - 28 Patients
    - Non-skull stripped
    - T1, T2, PD-weighted

Multimodal Synthesis
Training: Loss Function C1

- The loss function used to train this network is a combination of three terms:
  - C1
  - C2
  - C3

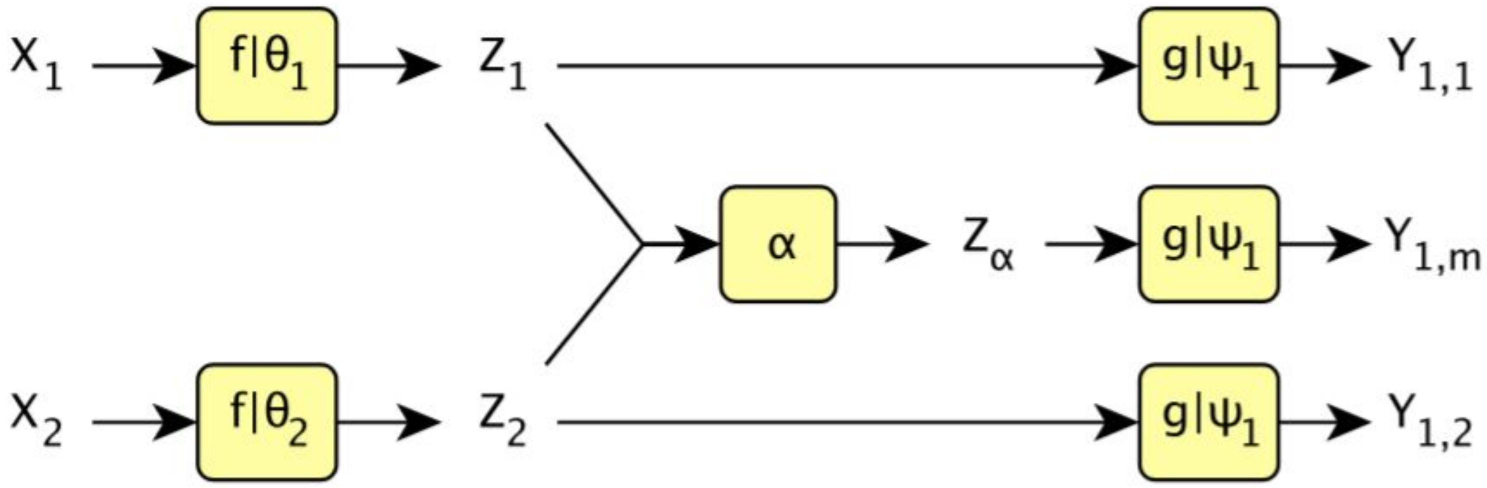
- Cost Component 1 (C1)

$$c_1(k|\boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m} MAE(g(f(X_i^k|\theta_i)|\psi_j), Y_j^k)$$

- Mean Absolute Error between each input and output.
- Each modality's individual latent representation should produce all outputs as accurately as possible.

Multimodal Synthesis

Training: Loss Function C1



Multimodal Synthesis
Training: Loss Function C2

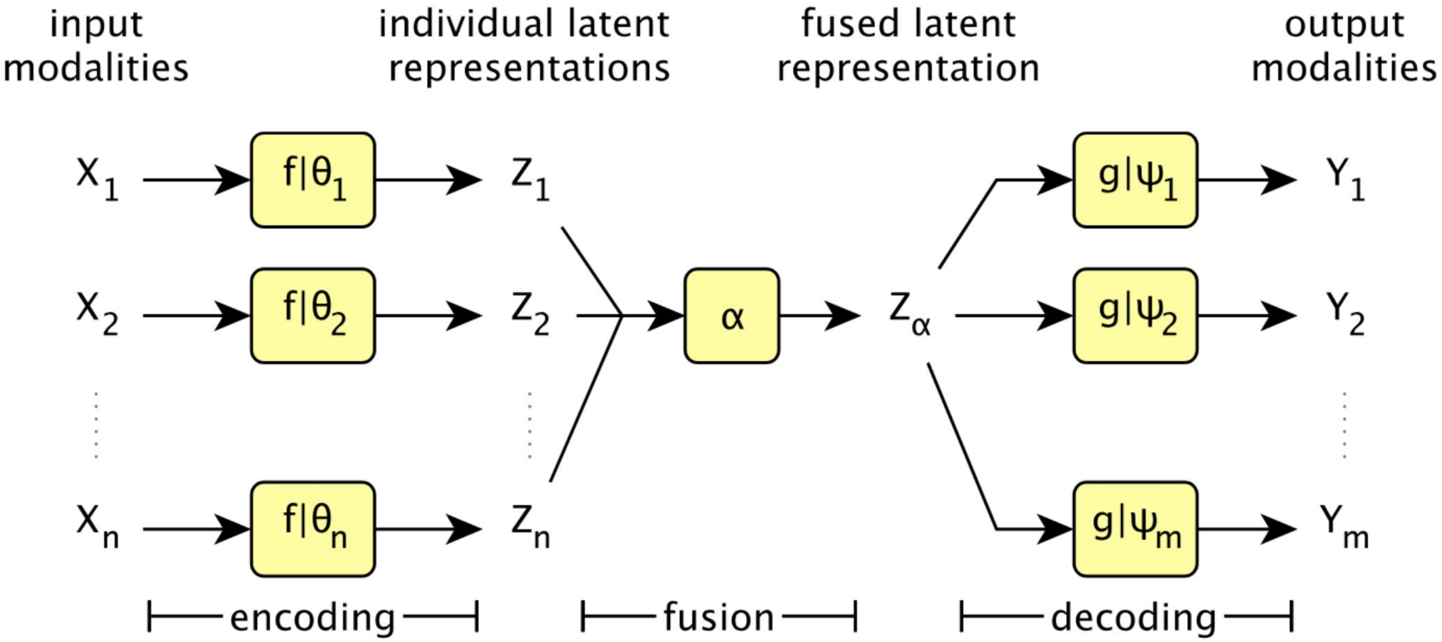
- Cost Component 2 (C2)

$$c_2(k|\boldsymbol{\theta}) = \frac{1}{|C||P|} \sum_{c \in C} \sum_{p \in P} var(f(X_1^k|\theta_1)_{p,c}, \ldots, f(X_n^k|\theta_n)_{p,c})$$

- Euclidean Distance between representations should be low.
- The latent representations from all input modalities should be close in the Euclidean sense.

Multimodal Synthesis

Training: Loss Function C2



Multimodal Synthesis
Training: Loss Function C3

- Cost Component 3 (C3)

$$c_3(k|\boldsymbol{\theta}, \boldsymbol{\psi}) \\ = \frac{1}{m}\sum_{j=1}^{m} MAE(g(\alpha(f(X_1^k|\theta_1), \ldots, f(X_n^k|\theta_n))|\psi_j), Y_j^k)$$

- Reconstruction error when using fused representation, using Mean Absolute Error.
- The latent representations from all input modalities should be close in the Euclidean sense.

Results

Evaluation Metrics

- Reconstruction of missing modality is evaluated using the following criteria
  - Mean Squared Error (MSE)
  - Peak Signal to Noise Ratio (PSNR)
  - Structured Similarity Index (SSIM)

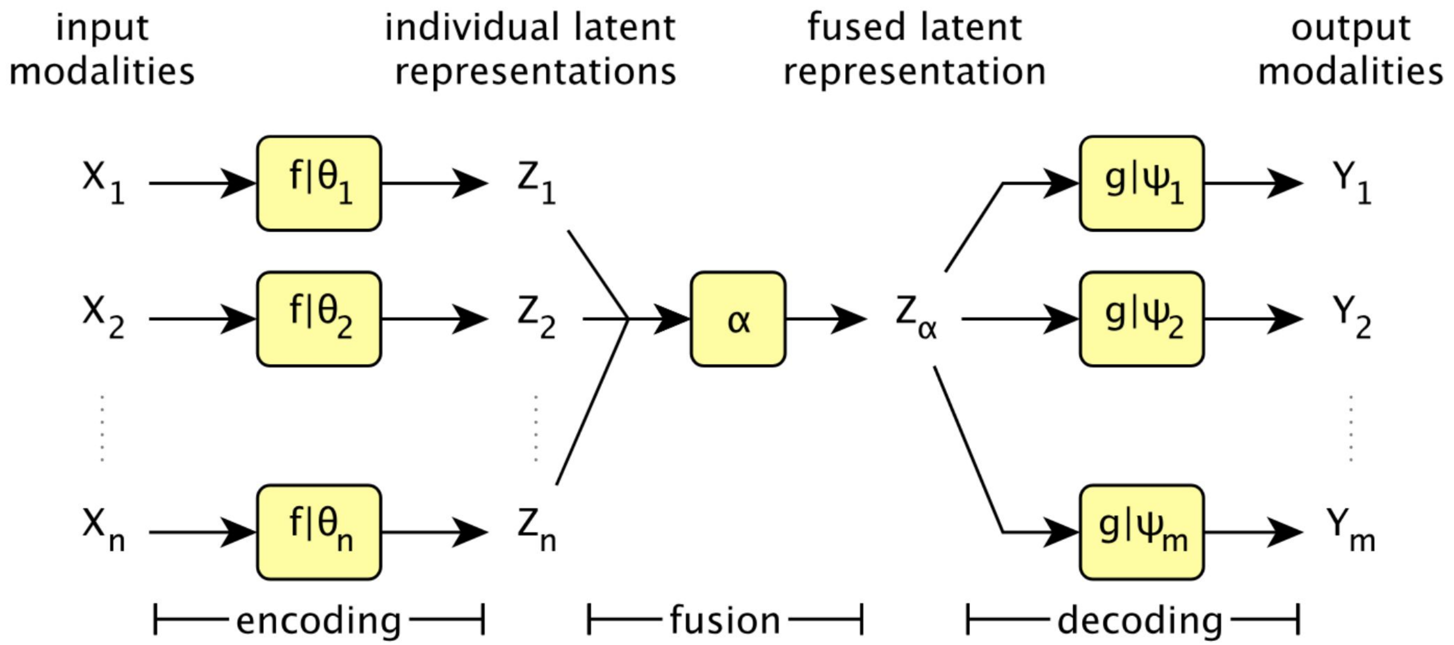
Results

Size of Latent Representations

TABLE I

COMPARISON OF DIFFERENT SIZED LATENT REPRESENTATIONS FOR T1, T2, DWI $\rightarrow$ FLAIR

	4 channels	8 channels	16 channels
MSE	0.184 (0.07)	0.191 (0.08)	**0.171** (0.06)
SSIM	0.866 (0.02)	0.865 (0.02)	0.869 (0.02)
PSNR	31.61 (1.69)	31.50 (1.72)	**31.10** (1.59)


input modalities
individual latent representations
fused latent representation
output modalities
X_1
$f|\theta_1$
Z_1
X_2
$f|\theta_2$
Z_2
X_n
$f|\theta_n$
Z_n
α
Z_α
$g|\psi_1$
Y_1
$g|\psi_2$
Y_2
$g|\psi_m$
Y_m
encoding
fusion
decoding

Results

Unimodal Models T1 -> T2 and T1 -> T2FLAIR

TABLE II

T1 → T2 AND T1 → FLAIR SYNTHESIS FROM UNIMODAL MODELS ON ISLES DATASET

T2	MP [2]	LSDN [7]	REPLICA [9]	Proposed
MSE	0.397 (0.15)	0.345 (0.12)	0.325 (0.12)	0.299 (0.11)
SSIM	0.798 (0.02)	0.811 (0.03)	0.823 (0.24)	**0.831** (0.03)
PSNR	25.22 (0.96)	25.22 (1.36)	25.51 (1.20)	25.78 (1.39)

FLAIR	MP [2]	LSDN [7]	REPLICA [9]	Proposed
MSE	0.343 (0.12)	0.286 (0.10)	0.301 (0.11)	0.268 (0.10)
SSIM	0.802 (0.03)	0.820 (0.03)	0.814 (0.03)	**0.831** (0.04)
PSNR	28.81 (2.13)	29.61 (2.17)	29.43 (2.25)	29.99 (2.24)

Results
Unimodal Models T1 -> T2 and T1 -> FLAIR on ISLES

TABLE II

T1 → T2 AND T1 → FLAIR SYNTHESIS FROM UNIMODAL MODELS ON ISLES DATASET

T2	MP [2]	LSDN [7]	REPLICA [9]	Proposed
MSE	0.397 (0.15)	0.345 (0.12)	0.325 (0.12)	0.299 (0.11)
SSIM	0.798 (0.02)	0.811 (0.03)	0.823 (0.24)	**0.831** (0.03)
PSNR	25.22 (0.96)	25.22 (1.36)	25.51 (1.20)	25.78 (1.39)
FLAIR	MP [2]	LSDN [7]	REPLICA [9]	Proposed
MSE	0.343 (0.12)	0.286 (0.10)	0.301 (0.11)	0.268 (0.10)
SSIM	0.802 (0.03)	0.820 (0.03)	0.814 (0.03)	**0.831** (0.04)
PSNR	28.81 (2.13)	29.61 (2.17)	29.43 (2.25)	29.99 (2.24)

Results

Unimodal Models T1 -> T2 and T1 -> FLAIR on BRATS

TABLE III

T1 → T2 AND T1 → FLAIR SYNTHESIS FROM UNIMODAL MODELS ON BRATS DATASET

T2	LSDN [7]	REPLICA [9]	Proposed
MSE	0.449 (0.12)	0.573 (0.17)	**0.333** (0.13)
SSIM	0.909 (0.02)	0.901 (0.01)	**0.929** (0.17)
PSNR	30.12 (1.62)	28.62 (1.69)	**30.96** (1.85)

FLAIR	LSDN [7]	REPLICA [9]	Proposed
MSE	0.332 (0.16)	0.432 (0.17)	**0.283** (0.14)
SSIM	0.887 (0.01)	0.870 (0.01)	**0.897** (0.01)
PSNR	29.68 (1.56)	28.32 (1.38)	**30.32** (1.61)

Results

Synthesize T2FLAIR from Multi-modal input data

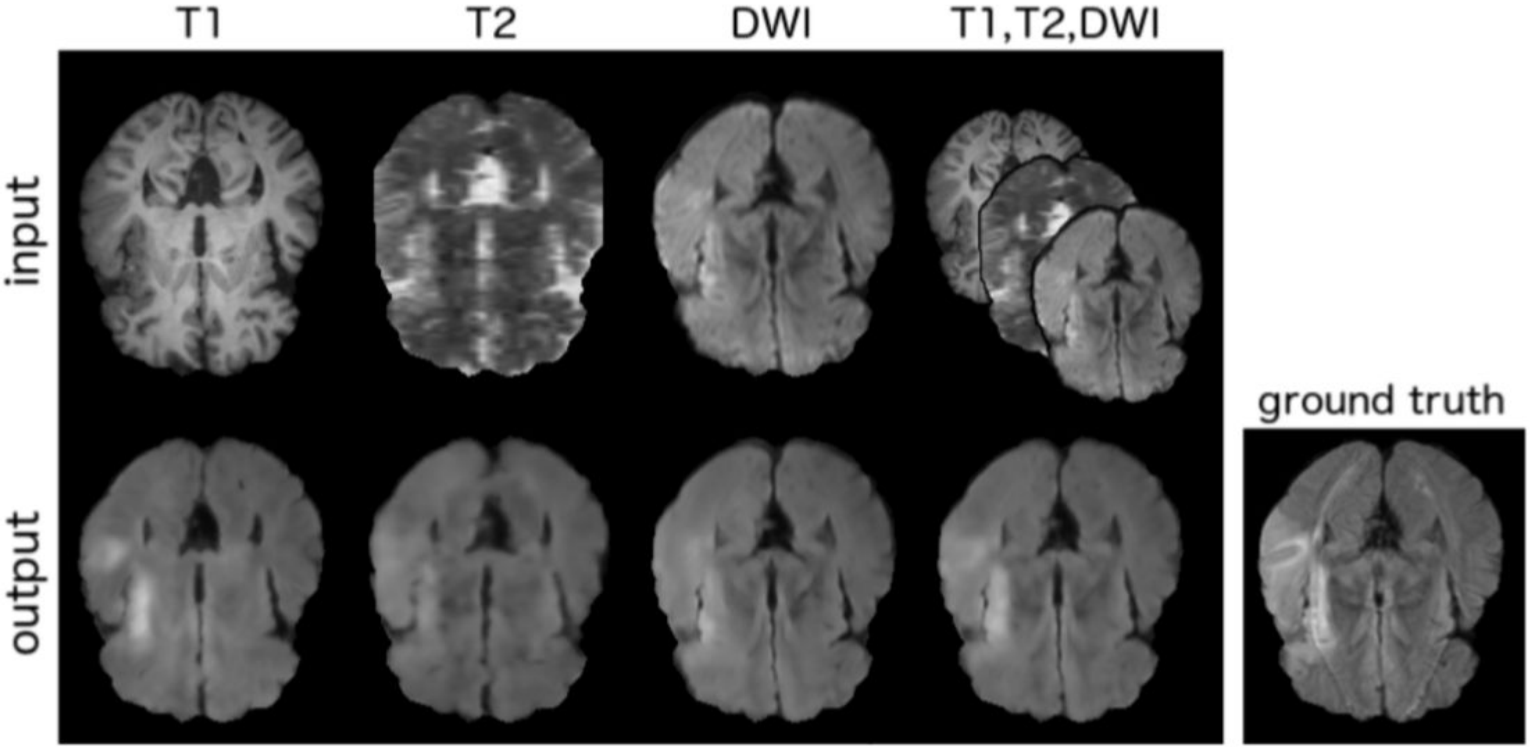
TABLE IV

SYNTHESIS OF FLAIR IMAGES WHEN TRAINING IN THE *Experiment A* AND *Experiment B* SETUPS

Combinations of Input			*MSE (FLAIR modality)*		
T1	*T2*	*DWI*	*REPLICA*	*Proposed: Exp. A*	*Proposed: Exp. B*
✓	—	—	0.301 (0.11)	**0.268** (0.10)	**0.249** (0.09)
—	✓	—	0.374 (0.16)	**0.328** (0.14)	**0.321** (0.12)
—	—	✓	0.278 (0.09)	0.303 (0.13)	0.285 (0.13)
—	✓	✓	0.235 (0.08)	0.215 (0.09)	0.214 (0.09)
✓	—	✓	0.225 (0.08)	**0.208** (0.09)	**0.198** (0.02)
✓	✓	—	0.271 (0.12)	**0.218** (0.08)	**0.214** (0.08)
✓	✓	✓	0.210 (0.08)	**0.171** (0.06)	**0.171** (0.06)
Average:			0.271	**0.244**	**0.236**

Results

Qualitative Results



Results
Synthesize Abnormal Data

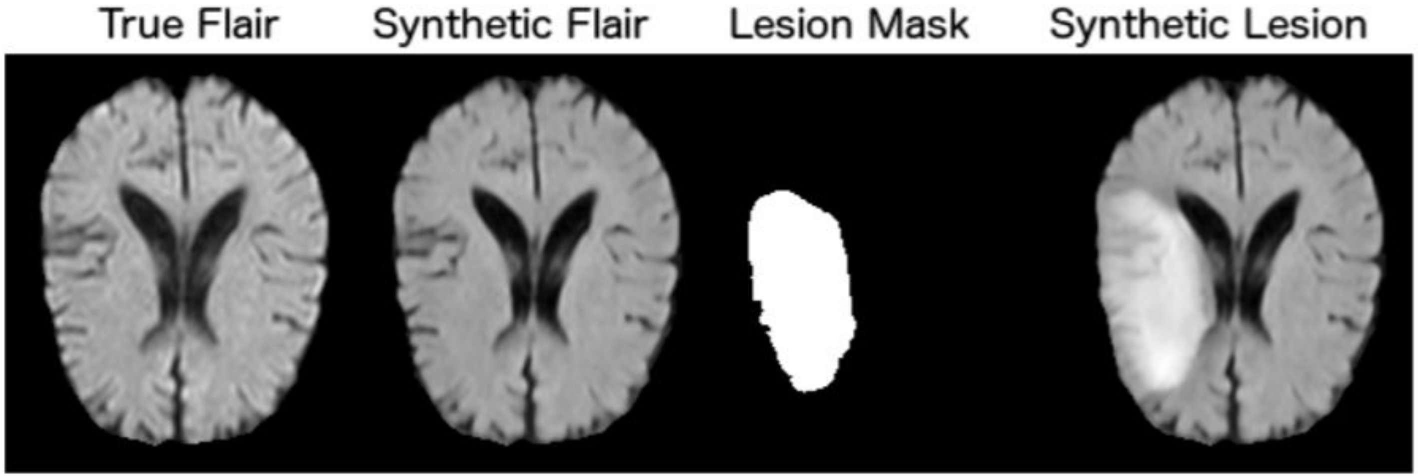


Fig. 10. Synthesis of a lesion by including a segmentation mask when synthesising an otherwise healthy image. This subject is taken from ISLES dataset in the FLAIR modality.

Results

Synthesize Different Views from Pre-Trained Model

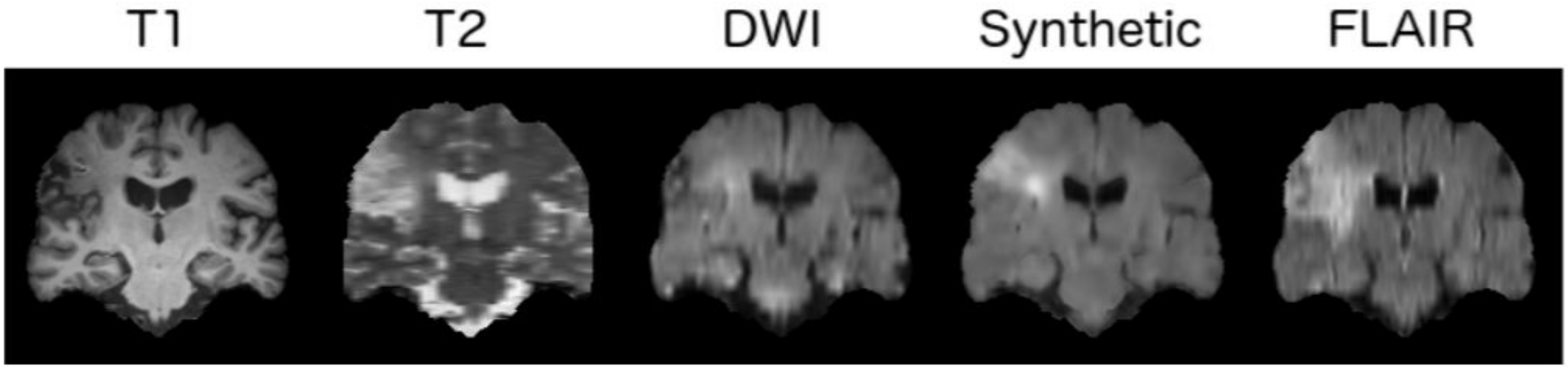


Fig. 11. A visual demonstration of robustness of our model to view transfer. We take the model trained on axial-plane slices and test using coronal-plane slices (shown). The image show the T1, T2 and DWI input slices, the synthesised FLAIR slice, and the ground-truth FLAIR image respectively.

Questions?

Thank You!