LARGO: Low-Rank Hypernetwork for Handling Missing Modalities

Niels Vyncke **Pooya Ashtari** **Aleksandra Pižurica**
Department of Telecommunications and Information Processing
Ghent University, Belgium
{niels.vyncke, pooya.ashtari, aleksandra.pizurica}@ugent.be

Abstract

Addressing missing modalities is an important challenge in multimodal image analysis and often relies on complex architectures that do not transfer easily to different datasets without architectural modifications or hyperparameter tuning. While most existing methods tackle this problem in *feature* space by engineering representations that are robust to missing inputs, we instead operate in *weight* space. We propose LARGO, a hypernetwork that compresses the $2^N - 1$ dedicated missing-modality models into a single network by modelling the convolutional weights using the Canonical Polyadic (CP) tensor decomposition. Extensive experimental validation on BraTS 2018 (4 modalities, 15 scenarios) and ISLES 2022 (3 modalities, 7 scenarios) shows that our method ranks first in 47 out of 52 configurations, achieving average Dice improvements of +0.68% and +2.53% over state-of-the-art baselines (mmFormer, M^{3}AE, ShaSpec, SimMLM). A proof-of-concept experiment on avMNIST suggests that LARGO may extend beyond medical imaging to heterogeneous non-medical modalities. Code is available at `https://github.com/nielsvyncke/LARGO`.

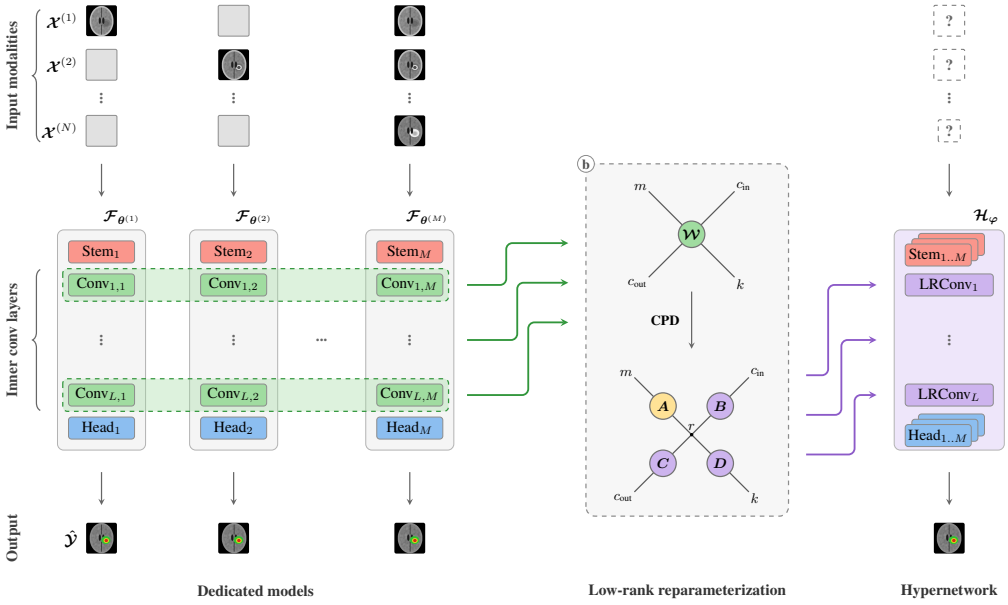
1 Introduction

Multimodal imaging combines complementary information from different acquisition protocols to improve accuracy [Baltrušaitis et al., 2018, Bayoudh et al., 2022]. In brain tumor segmentation, T2-weighted imaging reveals edema through fluid sensitivity, while contrast-enhanced T1-weighted imaging highlights active tumor regions through structural disruption [Baid et al., 2021, Menze et al., 2015, Bakas et al., 2017]. The synergy of multiple modalities enables precise delineation of tumor subregions that would be ambiguous from any single modality alone.

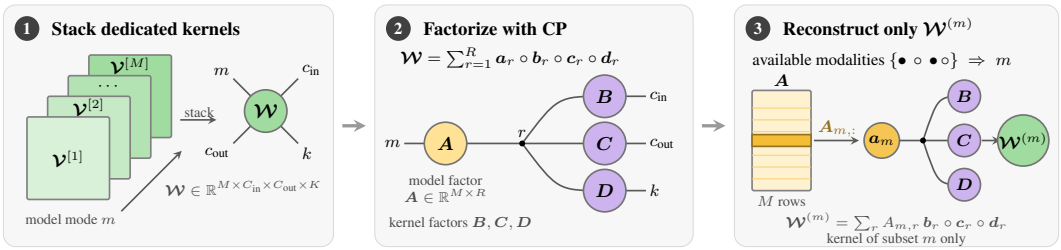
However, acquiring all modalities is not always feasible. Equipment availability, scan time constraints, patient conditions, and clinical protocols frequently result in incomplete modality sets [Wu et al., 2024, Azad et al., 2025]. Standard deep learning models cannot handle such missing inputs without explicit design considerations, creating a critical gap between research benchmarks—which assume complete data—and clinical deployment scenarios.

Existing approaches to missing modalities face a trade-off between *specialization*—optimal performance on a specific subset—and *scalability*—efficient support of all $2^N - 1$ combinations. Single-model methods such as zero-imputation [Pemberton et al., 2023] or shared-encoder architectures [Havaei et al., 2016, Dorent et al., 2019, Zhang et al., 2022] force one network to cover all combinations, compromising specific configurations; generative approaches [Chartsias et al., 2018, Dar et al., 2019] risk hallucinating clinically dangerous features; and knowledge distillation [Maheshwari et al., 2024, Wang et al., 2023b] requires multi-stage training. Training dedicated models per combination maximizes specialization but scales exponentially in storage and compute.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).



(a) Network level: the M dedicated models are compressed into one hypernetwork.



(b) Layer level: the low-rank reparameterization block of (a) for one convolutional layer.

Figure 1: Illustration of LARGO. (a) The dedicated models $\mathcal{F}_{\boldsymbol{\theta}^{(m)}}$ for each missing-modality case $m = 1, ..., M$ are compressed into a single hypernetwork $\mathcal{H}_{\boldsymbol{\varphi}}$, by representing the kernel weights $\boldsymbol{\mathcal{W}}$ of their convolutional layers using CPD decomposition, which we refer to as low-rank convolutional layers (LRConv). The stem and head layers remain uncompressed. (b) For each such layer, (1) the kernels of the $M = 2^N - 1$ dedicated models are stacked along a new *model* mode into $\boldsymbol{\mathcal{W}}$; (2) $\boldsymbol{\mathcal{W}}$ is factorized with the CP decomposition into a model factor $\boldsymbol{A}$ and kernel factors $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$, trained jointly from scratch; (3) at inference, the available subset m selects row $\boldsymbol{A}_{m,:}$, and only the kernel $\boldsymbol{\mathcal{W}}^{(m)}$ of that subset is reconstructed.

We propose a different approach that achieves *both* specialization and scalability. Our key observation is that dedicated models for different modality combinations are not independent: models sharing access to the same modalities must learn similar patterns, implying substantial parameter redundancy across the family. We exploit this redundancy through tensor decomposition.

Our approach, LARGO (**L**ow-r**A**nk hype**R**network for missin**G** m**O**dalities), constructs a hypernetwork by stacking the convolutional weights of all $2^N - 1$ dedicated models along a new *model* dimension and applying Canonical Polyadic (CP) decomposition. The entire family is represented through compact factor matrices, with one factor indexing modality combinations, yielding a single network that implicitly contains all dedicated models at the parameter budget of just one.

Unlike feature-space methods, LARGO operates in *weight* space and is the first to represent the exponential family of $2^N - 1$ missing-modality models within a single compressed weight manifold—a formulation absent from both prior tensor-decomposition compression of individual networks [Ashtari et al., 2020, Kim et al., 2015, Lebedev et al., 2015] and hypernetwork literature [Ha et al., 2016, Zhang et al., 2018]. It does not rely on auxiliary losses, teacher networks, or multi-stage training. The result is a jointly optimized low-rank manifold of $2^N - 1$ subset-specific networks (Figure 1b).

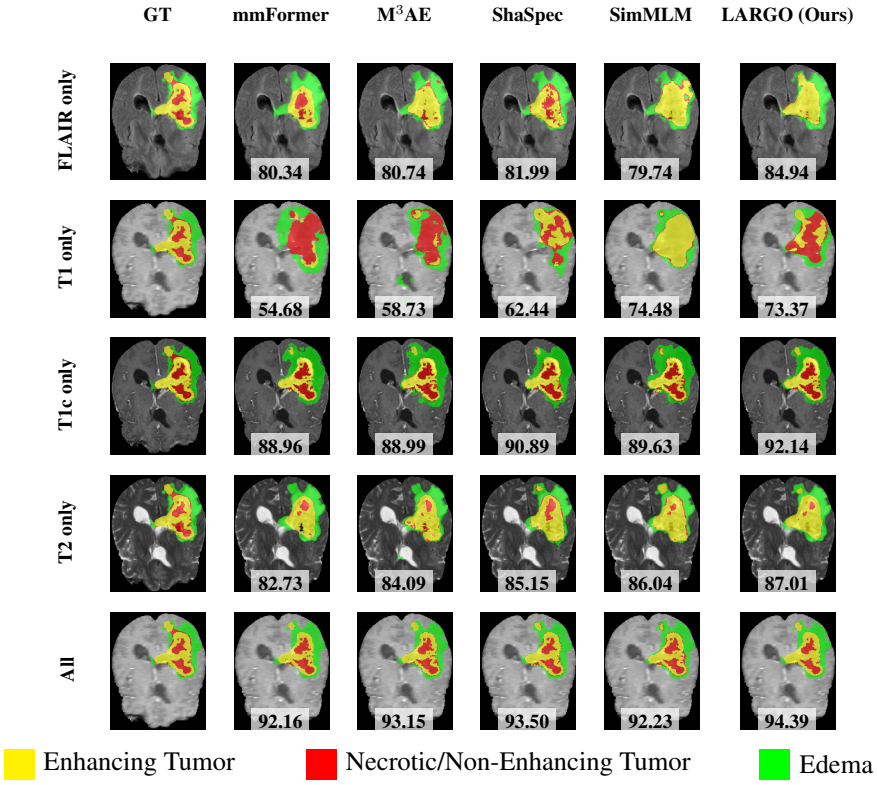


Figure 2: Qualitative comparison on BraTS 2018. Ground truth (GT) and predictions from different methods under single-modality and all-modality settings are shown. Yellow indicates enhancing tumor, red indicates the necrotic and non-enhancing tumor regions, and green indicates edema. Dice scores are reported for each prediction.

Our contributions are:

- A novel hypernetwork formulation that jointly reparameterizes the kernel weights of the entire family of $2^N - 1$ dedicated missing-modality models into a single parameter-efficient representation using canonical polyadic decomposition.
- State-of-the-art results on BraTS 2018 and ISLES 2022: LARGO ranks first in 47 out of 52 missing-modality scenarios, with +0.68% and +2.53% average Dice gains over mmFormer, M³AE, ShaSpec, and SimMLM.
- A proof-of-concept on avMNIST (image+audio) showing that LARGO extends to non-medical, heterogeneous-modality classification, with +0.44% average accuracy over the baselines.

Notation. Vectors are denoted by bold lowercase (e.g., $\boldsymbol{x}$), matrices by bold uppercase (e.g., $\boldsymbol{U}$), tensors by calligraphic bold uppercase (e.g., $\boldsymbol{\mathcal{W}}$), and sets by blackboard bold uppercase (e.g., $\mathbb{X}$). Individual elements are indexed by subscripts in non-bold notation: x_i, U_{ij}, $\mathcal{W}_{ijk}$.

2 Related work

We review prior work along three directions: incomplete multimodal learning, brain tumor segmentation under missing modalities, and hypernetworks with tensor-based network compression. Existing missing-modality methods mainly learn robust feature representations, impute missing inputs, or distill knowledge from complete-modality models [Wu et al., 2024, Azad et al., 2025]. In contrast, LARGO operates in *weight* space, jointly compressing the family of dedicated models associated with all $2^N - 1$ non-empty modality subsets through a shared low-rank parameterization.

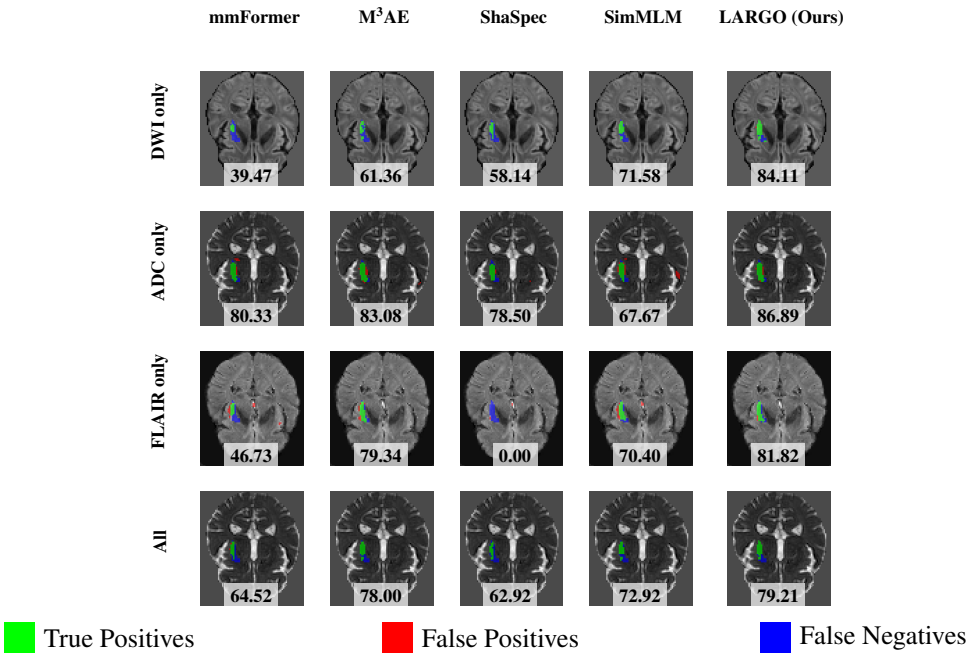


Figure 3: Qualitative comparison on the ISLES 2022 dataset. Predictions from different methods under single-modality and all-modality settings are shown. Green denotes true positives, red false positives, and blue false negatives. Dice scores (%) are reported for each example.

Single-model approaches with modality dropout. A widely used strategy replaces missing modalities with zeros, often combined with sparsified training that randomly drops modalities so that one network can process arbitrary subsets at test time [Pemberton et al., 2023]. While simple, this forces a single fixed parameter set to serve all $2^N - 1$ combinations and inevitably compromises configurations that would benefit from different fusion behavior.

Knowledge transfer, generative, and dedicated models. Knowledge distillation [Maheshwari et al., 2024, Wang et al., 2023b] trains incomplete-modality students to mimic full-modality teachers, while SMIL [Ma et al., 2021] casts the problem as Bayesian meta-learning. Generative approaches [Chartsias et al., 2018, Dar et al., 2019] synthesize missing modalities but risk hallucinating clinically dangerous features and require multi-stage training. Training a separate dedicated model per non-empty subset maximizes specialization but scales exponentially in storage and compute [Wu et al., 2024]; LARGO aims at retaining this specialization while replacing explicit storage by a jointly optimized low-rank parameterization.

Recent state of the art. LS3M [Zhang et al., 2025b] reorders modality feature sequences with differentiable permutations and fuses them with state-space blocks; Hierarchical Gradient Alignment (HGA) [Shi et al., 2026] aligns gradients across configurations to handle realistic, imbalanced incomplete-modality distributions; and UniME [Song et al., 2026] pretrains a ViT uni-encoder with masked image modelling and adds modality-specific CNN encoders before fusion. All three act on features or on the optimizer and keep the model representation, whereas LARGO changes the parameterization of the model family and could be combined with them. We do not claim empirical superiority over them: UniME is evaluated on BraTS 2023/2024 and HGA uses a different, realistic-incompleteness protocol.

Brain tumor segmentation under missing modalities. Many methods improve robustness by combining modality-specific encoders with feature aggregation. HeMIS [Havaei et al., 2016] processes each modality independently and aggregates feature statistics, UHVED [Dorent et al., 2019] maps modalities into a shared variational latent space, mmFormer [Zhang et al., 2022] uses transformer-based cross-modal attention, and ShaSpec [Wang et al., 2023a] separates shared and modality-specific representations through dedicated branches. M³AE [Liu et al., 2023] and SimMLM [Li et al., 2025]

likewise fuse in feature space; we compare against mmFormer, M³AE, ShaSpec, and SimMLM in Section 4. All of these operate in *feature* space, whereas LARGO acts in *weight* space.

Hypernetworks and dynamic weight generation. Hypernetworks generate or modulate target-network weights conditioned on an input, task, or architecture descriptor [Ha et al., 2016, Jia et al., 2016, Zhang et al., 2018, Nava et al., 2022]. LARGO is related in spirit, since it reconstructs subset-specific weights, but does not learn an auxiliary generator: it stores the entire missing-modality model family directly in a factorized tensor, indexing modality subsets through one factor matrix and reconstructing convolutional weights by tensor contraction.

Tensor decomposition for multimodal fusion. Tensor factorization has also been used to model high-order cross-modal interactions, e.g. in tensor fusion networks, low-rank multimodal fusion, and multi-way multimodal attention [Zadeh et al., 2017, Liu et al., 2018, Zhang et al., 2025a]. These methods factorize *feature* representations or fusion operators; LARGO instead factorizes *model weights* across modality-subset-specific networks.

Tensor decomposition for network compression. Tensor decompositions [Sidiropoulos et al., 2017, Kolda and Bader, 2009] compress neural networks by approximating high-order weight tensors with low-rank factors. CP decomposition has been applied to convolutional layers to reduce parameters and computation [Kim et al., 2015, Lebedev et al., 2015], and Ashtari et al. [2020] compressed a single U-Net for brain tumor segmentation. These methods perform *within-model* compression—factorizing the weights of one network. LARGO instead performs *inter-model* compression: it introduces a model dimension corresponding to modality subsets and jointly compresses the entire family of $2^N - 1$ models—to our knowledge, the first use of tensor decomposition to jointly parameterize the family of dedicated missing-modality models. Moreover, within-model compression decomposes one pretrained network and fine-tunes it, whereas LARGO learns all subset-specific networks jointly from scratch, so the shared factors must represent models receiving overlapping or disjoint modality subsets at once.

3 Method

3.1 Problem formulation

We define the base network $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta}}$, with parameters $\boldsymbol{\theta}$, that takes all available modalities $\{\boldsymbol{\mathcal{X}}^{(n)}\}_{n=1}^{N}$ as input. Ideally, we want to train different models on non-empty subsets of the available modalities, such as $\{\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta}^{(m)}}\}_{m=1}^{M}$, where $M = 2^N - 1$.

Instead of training M copies of $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta}}$, we want to optimize this by applying a dimensionality constraint on the set of networks, such that the total number of parameters remains the same as for $\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta}}$. Therefore, our proposed hypernetwork LARGO is designed to approximate these models, i.e. $\boldsymbol{\mathcal{H}}_{\boldsymbol{\varphi}} \approx \{\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta}^{(m)}}\}_{m=1}^{M}$, which is fundamentally different from the way existing approaches handle missing modalities.

3.2 Low-rank hypernetwork architecture

3.2.1 Hypernetwork design

We compress the networks by applying compression across the corresponding (transposed) convolutional layers.

Stem and head layers. No compression is applied. We retain M copies of this layer. For the stem layers, we adapt the number of input channels to the number of input modalities. This does not significantly affect the compression ratio, as these layers are very small due to the limited number of input and output channels. To obtain a compression ratio of $1/M$, we would have to set the rank to 1, which severely limits the expressive capacity of the model and leads to poor representations.

Inner layers. We compress the parameters across all M dedicated models using tensor decomposition. We represent the weight tensor for the l-th layer as $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{M \times C_{\text{in}} \times C_{\text{out}} \times K}$, where C_{out} and

C_{in} are the output and input channels, and K represents the flattened spatial kernel dimensions. We suppress the index l in the remainder of this section for readability.

3.2.2 Low-rank compression of the convolutional layers

Convolutional layer. In any convolutional neural network, the output neuron at location (h, w) in the channel c_{out} is computed as:

$$\boldsymbol{\mathcal{Y}}_{c_{\text{out}},h,w} = (\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{V}})_{c_{\text{out}},h,w} = \left\langle \boldsymbol{\mathcal{S}}_{h,w}, \boldsymbol{\mathcal{V}}^{(c_{\text{out}})} \right\rangle + \mathbf{b}_{c_{\text{out}}}, \tag{1}$$

where $\boldsymbol{\mathcal{V}} = \left\{\boldsymbol{\mathcal{V}}^{(c_{\text{out}})}\right\}_{c_{\text{out}}=1}^{C_{\text{out}}}$, $\boldsymbol{\mathcal{V}}^{(c_{\text{out}})} \in \mathbb{R}^{C_{\text{in}} \times K_h \times K_w}$ represents the tensor containing the kernel weights, with $\mathbf{b} \in \mathbb{R}^{C_{\text{out}}}$ the bias vector, and where $\boldsymbol{\mathcal{S}}_{h,w}$ is the sliding window operator on the input tensor $\boldsymbol{\mathcal{X}}$ defined as $\boldsymbol{\mathcal{S}}_{h,w} \triangleq \boldsymbol{\mathcal{X}}_{h-\Delta_h:h+\Delta_h, w-\Delta_w:w+\Delta_w}$, with $K_h = 2\Delta_h + 1$ and $K_w = 2\Delta_w + 1$.

Low-rank decomposition. We decompose the weight tensor $\boldsymbol{\mathcal{W}} = \left\{\boldsymbol{\mathcal{V}}^{[m]}\right\}_{m=1}^{M}$ containing the kernel weights of the M dedicated models, using Low-rank decomposition along the modes of the tensor (M, C_{in}, C_{out}, and K). For example, the Canonical Polyadic (CP) decomposition has the following structure:

$$\boldsymbol{\mathcal{W}} = \sum_{r=1}^{R} \boldsymbol{a}_r \circ \boldsymbol{b}_r \circ \boldsymbol{c}_r \circ \boldsymbol{d}_r \tag{2}$$

where $\boldsymbol{a}_r \in \mathbb{R}^M$, $\boldsymbol{b}_r \in \mathbb{R}^{C_{\text{in}}}$, $\boldsymbol{c}_r \in \mathbb{R}^{C_{\text{out}}}$, and $\boldsymbol{d}_r \in \mathbb{R}^K$ are the r-th columns of factor matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ respectively.

This decomposition is visually represented as a tensor network diagram in Figure 1b, which also summarizes the three stages of the construction: stacking the dedicated kernels, factorizing the joint tensor, and reconstructing the kernel of the target subset. In the supplementary material, we present an ablation study comparing CP and Tucker decomposition to assess the effect of the choice of decomposition.

Rank selection. The rank R is chosen to maintain the number of parameters per layer, thereby applying a compression of $1/M$. Each of the original convolutional layers has $C_{\text{in}}C_{\text{out}}K$ parameters, while the low rank convolutional layer has $(M + C_{\text{in}} + C_{\text{out}} + K)R$ parameters. Thus, we set the rank for the $1/M$ compression as follows:

$$R = \left\lfloor \frac{C_{\text{in}}C_{\text{out}}K}{M + C_{\text{in}} + C_{\text{out}} + K} \right\rfloor \tag{3}$$

In the supplementary material, we present an ablation study on the rank R, comparing different values to assess the effect of varying compression ratios.

Bias. We retain unique bias terms for all M layers, resulting in $C_{\text{out}} \times M$ parameters. This has only a small effect on the total number of parameters.

Initialization. The weight matrices $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ representing the input channels, output channels, and flattened spatial kernel dimensions, respectively, are initialized using Kaiming normal initialization. The weight matrix $\boldsymbol{A}$, representing the number of models, is initialized with ones, and the bias weights with zeros. Note that weight matrices are trained from scratch, so we do not compute the CPD from pretrained weights.

Normalization. To enforce numerical stability and improve gradient flow, we normalize the atoms corresponding to the input channels, output channels, and kernel dimensions. The atom containing M is not normalized. This does not affect the low-rank representations, because the norms of these three atoms can be absorbed into the other atom.

Table 1: Dice scores (%) ↑ on BraTS 2018 (5-fold cross-validation; mmF: mmFormer, ShaS: ShaSpec, SimM: SimMLM). •/○: modality present/absent. Best and second-best scores per tumor region in **bold** and underlined.

Fl	T1	T1c	T2	Whole tumor (WT)					Tumor core (TC)					Enhancing tumor (ET)				
				mmF	M³AE	ShaS	SimM	LARGO (Ours)	mmF	M³AE	ShaS	SimM	LARGO (Ours)	mmF	M³AE	ShaS	SimM	LARGO (Ours)
•	○	○	○	86.71	87.28	87.50	88.20	**88.85**	66.73	67.82	71.15	68.27	**72.28**	35.97	42.09	43.27	42.33	**43.74**
○	•	○	○	68.38	74.32	77.09	78.06	**78.41**	60.20	64.54	70.15	68.02	**70.51**	29.41	36.84	42.12	40.02	**42.36**
○	○	•	○	74.95	74.49	77.32	76.05	**79.13**	80.71	80.54	82.77	81.97	**83.70**	73.87	74.34	75.07	75.88	**76.21**
○	○	○	•	83.26	84.43	84.70	85.17	**86.20**	66.31	68.44	71.88	69.89	**72.44**	37.91	42.59	46.35	44.76	**46.65**
•	•	○	○	88.35	88.82	88.92	88.53	**89.84**	71.80	71.51	**75.39**	70.19	75.11	40.67	44.18	46.47	43.14	**47.61**
•	○	•	○	88.89	89.28	89.04	88.70	**90.09**	82.81	84.06	83.85	82.63	**84.56**	75.11	**76.31**	75.76	75.99	76.03
•	○	○	•	88.82	89.32	89.12	89.30	**89.99**	71.40	71.62	74.50	71.56	**74.99**	42.60	45.27	47.59	46.04	**48.83**
○	•	•	○	78.07	76.83	79.90	79.35	**80.45**	82.19	82.28	83.69	82.50	**84.10**	75.73	74.59	76.29	75.89	**76.82**
○	•	○	•	85.13	85.88	86.27	85.58	**86.88**	70.61	71.03	**74.37**	71.30	73.85	41.88	44.63	47.76	45.66	**48.11**
○	○	•	•	86.16	85.60	86.23	85.83	**87.20**	83.47	83.87	84.15	82.61	**84.80**	76.62	76.23	75.93	76.01	**76.82**
•	•	•	○	89.32	89.39	89.35	88.82	**90.22**	83.46	**84.41**	84.26	82.77	84.37	76.41	76.51	77.05	75.98	**77.35**
•	•	○	•	89.27	89.52	89.52	89.35	**90.14**	73.04	72.89	75.94	72.45	**76.09**	45.57	46.00	48.80	46.62	**49.22**
•	○	•	•	89.80	90.01	89.73	89.46	**90.49**	83.60	84.25	84.13	82.71	**84.32**	76.03	76.74	75.57	76.03	**77.42**
○	•	•	•	86.51	85.80	86.66	85.86	**87.47**	83.48	84.05	84.50	82.78	**84.73**	76.75	76.48	76.82	75.99	**77.09**
•	•	•	•	89.93	90.09	89.88	89.44	**90.44**	83.77	**84.69**	84.40	82.73	84.47	76.13	76.59	76.40	76.02	**78.18**
Average				84.90	85.40	86.08	85.85	**87.05**	76.24	77.07	79.01	76.82	**79.35**	58.71	60.63	62.08	61.09	**62.83**

Forward pass and modality selection. During inference, given a specific modality combination m, the hypernetwork reconstructs the dedicated weights corresponding to m. For CP decomposition, the forward pass computes:

$$\boldsymbol{\mathcal{W}}^{(m)} = \sum_{r=1}^{R} A_{m,r} \left(\boldsymbol{b}_r \circ \boldsymbol{c}_r \circ \boldsymbol{d}_r \right) \tag{4}$$

where $A_{m,r}$ is a scalar value at position m in the rank vector $\boldsymbol{a}_r$. This is equivalent to computing the whole joint kernel weights tensor $\boldsymbol{\mathcal{W}}$, and then selecting the kernels for model m, but avoids the computational overhead of computing unused parameters. The construction of the hypernetwork is visually represented in Figure 1a.

3.3 Training procedure

Loss function. We use the standard segmentation loss $\mathcal{L}_{\text{seg}} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{CE}}$. We train the hypernetwork with random modality dropout combined with full-modality guidance: at each iteration, we sample a modality subset $m \sim \mathcal{U}(1, M)$ and minimize

$$\mathcal{L}_{\text{total}} = \tfrac{1}{2}\big(\mathcal{L}_{\text{seg}}(\boldsymbol{\mathcal{H}}_{\boldsymbol{\varphi}}(\boldsymbol{\mathcal{X}}_m; m), \boldsymbol{\mathcal{Y}}) + \mathcal{L}_{\text{seg}}(\boldsymbol{\mathcal{H}}_{\boldsymbol{\varphi}}(\boldsymbol{\mathcal{X}}; M), \boldsymbol{\mathcal{Y}})\big), \tag{5}$$

where $\boldsymbol{\mathcal{X}}_m$ is the input with modality combination m, $\boldsymbol{\mathcal{X}}$ is the full-modality input, and $\boldsymbol{\mathcal{Y}}$ the target. The first term ensures the hypernetwork handles all $2^N - 1$ missing-modality scenarios; the second term provides full-modality guidance throughout training.

4 Experiments

4.1 Experimental setup

Baselines. We compare our proposed hypernetwork LARGO against four state-of-the-art missing modality methods: mmFormer [Zhang et al., 2022], M³AE [Liu et al., 2023], ShaSpec [Wang et al., 2023a], and SimMLM [Li et al., 2025]. To ensure fair comparison, all models are trained from scratch under identical conditions: the same folds, learning rate, batch size, scheduler, optimizer, preprocessing, augmentation, patch size, cross-

Table 2: Model complexity on BraTS 2018, measured on the same hardware with the full-modality input. Best and second-best scores in **bold** and underlined.

Model	# Params (M)	Time per epoch	GFLOPs	Avg Dice (%) ↑	Speed (vol/s) ↑
mmFormer	36.7	2m 30s	138.8	73.28 ± 2.31	0.89
M³AE	2.1	1m 07s	333.7	74.37 ± 0.99	1.02
ShaSpec	187.7	2m 42s	1411.9	75.73 ± 0.89	0.68
SimMLM	7.8	1m 12s	242.1	74.59 ± 0.80	**1.14**
LARGO (Ours)	22.6	1m 51s	1155.1	**76.41** ± 0.78	**1.14**

validation protocol, and evaluation procedure, without method-specific hyperparameter tuning, so the reported differences isolate architectural contributions.

Architecture. We use nnUnet [Isensee et al., 2021] as base architecture for LARGO, which automatically configures the U-Net topology based on dataset properties. For avMNIST, we use a different architecture. Details of the model architecture are provided in the supplementary material.

Statistical analysis. Each patient contributes one paired observation (Dice averaged over all modality subsets and, for BraTS, tumor regions); we apply two-sided paired Wilcoxon signed-rank tests of LARGO against each baseline with Benjamini–Hochberg correction (Table 10, supplementary material). All patient-level comparisons remain significant, as do 57 of 60 per-subset comparisons on BraTS 2018 and 26 of 28 on ISLES 2022.

4.2 BraTS 2018 dataset

Dataset. The BraTS 2018 dataset [Baid et al., 2021, Menze et al., 2015, Bakas et al., 2017] contains 285 brain MRI cases with four modalities: FLAIR (fluid-attenuated inversion recovery), T1-weighted, T1-weighted with contrast enhancement (T1c), and T2-weighted. The task is to segment three tumor regions: whole tumor (WT), tumor core (TC), and enhancing tumor (ET). With four modalities, there are 15 possible missing-modality scenarios.

Training. All models are trained for 750 epochs using stochastic gradient descent with Nesterov momentum (0.99), learning rate (5×10^{-4}), and weight decay (10^{-5}).

Quantitative results. Table 1 presents comprehensive Dice scores across all 15 missing-modality combinations. LARGO consistently achieves the best performance, ranking first in all 15 scenarios of WT, 11 out of 15 for TC, and 14 out of 15 for ET, and second in all other cases. On average, LARGO achieves 87.05% for WT, 79.35% for TC, and 62.83% for ET, outperforming all baselines. Notably, ShaSpec achieves competitive results but requires 8.3× more parameters (187.7M vs. 22.6M). The largest gains occur in the difficult single-modality cases (e.g., T1-only), indicating that the shared factors preserve modality-specific behavior rather than collapsing to one generic representation; all gains are significant at the patient level (Table 10).

Table 3: Dice scores (%) ↑ on ISLES 2022 (5-fold cross-validation). •/○: modality present/absent. Best and second-best scores in **bold** and underlined.

DWI	ADC	FLAIR	Stroke lesion				
			mmFormer	M^{3}AE	ShaSpec	SimMLM	LARGO (Ours)
•	○	○	66.39	72.20	62.76	<u>74.11</u>	**74.23**
○	•	○	<u>40.03</u>	37.79	24.47	39.92	**42.49**
○	○	•	35.05	<u>38.60</u>	32.63	38.28	**41.09**
•	•	○	69.33	74.10	65.10	<u>74.39</u>	**74.67**
•	○	•	67.12	72.88	64.50	<u>74.54</u>	**75.75**
○	•	•	50.50	<u>50.82</u>	41.54	44.51	**53.70**
•	•	•	69.76	<u>74.62</u>	65.44	74.42	**76.82**
Average			56.88	<u>60.14</u>	50.92	60.02	**62.68**

Table 2 compares model complexity and efficiency. LARGO achieves the highest average Dice score (76.41%) while maintaining competitive inference speed (1.14 vol/s) and moderate parameter count (22.6M). LARGO is not the smallest or fastest model in absolute terms: mmFormer and M³AE use narrow convolutions and low-resolution transformer modules, whereas LARGO retains a full-width nnU-Net. Its compression claim is relative to the family it replaces: 15 dedicated nnU-Nets need about 338.6M parameters, whereas LARGO represents all of them with 22.60M, the budget of a single nnU-Net (22.57M), a ∼15× reduction. It also trains in a single stage, with an epoch time below mmFormer and ShaSpec.

Table 4: Model complexity on ISLES 2022. Best and second-best scores in **bold** and underlined.

Model	# Params (M)	Avg Dice (%) ↑	Speed (vol/s) ↑
mmFormer	35.3	56.88 ± 4.00	1.28
M³AE	21.7	<u>60.14</u> ± 3.13	**4.76**
ShaSpec	151.7	50.92 ± 2.51	1.56
SimMLM	5.9	60.02 ± 1.99	3.56
LARGO (Ours)	22.4	**62.67** ± 2.28	<u>3.88</u>

Qualitative results. Figure 2 visualizes segmentation results for single-modality scenarios. Yellow indicates enhancing tumor, red indicates necrotic/non-enhancing tumor, and green indicates edema. LARGO produces more accurate segmentations with fewer false positives and false negatives compared to all baselines, particularly in challenging single-modality cases with limited information.

4.3 ISLES 2022 dataset

Dataset. The ISLES 2022 dataset [de la Rosa et al., 2025, Hernandez Petzsche et al., 2022] contains 250 stroke MRI cases with three modalities: DWI (diffusion-weighted imaging), ADC (apparent diffusion coefficient), and FLAIR (fluid-attenuated inversion recovery). The task is to segment acute and sub-acute ischemic stroke lesions. With three modalities, there are 7 possible missing-modality scenarios.

Training. All models are trained for 500 epochs using stochastic gradient descent with Nesterov momentum (0.99), learning rate (5×10^{-4}), and weight decay (10^{-5}).

Quantitative results. Table 3 presents Dice scores across all 7 modality combinations. LARGO scores best in all 7 scenarios, with an average Dice score of 62.67%, outperforming M³AE (60.14%), SimMLM (60.02%), mmFormer (56.88%), and ShaSpec (50.92%). The improvement is particularly pronounced in challenging scenarios with limited modalities, such as ADC-only (42.49% vs. 40.03%) and FLAIR-only (41.09% vs. 38.60%). All gains are significant at the patient level (Table 10).

Table 4 compares model complexity and efficiency. LARGO achieves the highest average Dice score (62.67%) with 22.4M parameters and competitive inference speed (3.88 vol/s). While M³AE is faster (4.76 vol/s), it achieves lower accuracy (60.14%). This demonstrates our method's effectiveness on stroke lesion segmentation across diverse missing-modality scenarios.

Table 5: Accuracy (%) ↑ on avMNIST (predefined train-test split). ●/○: modality present/absent. Best and second-best scores in **bold** and underlined.

Image	Audio	Classification accuracy				
		mmFormer	M³AE	ShaSpec	SimMLM	LARGO (Ours)
●	○	94.67	93.33	94.00	94.00	**96.00**
○	●	**95.67**	95.00	95.00	94.67	**95.67**
●	●	**99.33**	**99.33**	**99.33**	**99.33**	**99.33**
Average		96.56	95.89	96.11	96.00	**97.00**

Qualitative results. Figure 3 visualizes segmentation results for single-modality scenarios. True positives appear in green, false positives in red, and false negatives in blue. LARGO consistently produces more accurate lesion boundaries and reduces false predictions compared to all baselines, demonstrating robust performance even with severely limited input information.

4.4 avMNIST dataset

As a proof-of-concept, we evaluate LARGO on a non-medical classification task with heterogeneous modalities.

Dataset. The audiovisual MNIST (avMNIST dataset)[1] provides 3000 samples (2100 train + 900 test) with 2 modalities: images and audio. The images are taken from the MNIST dataset, while audio samples are recorded from 6 different speakers (50 per digit).

Table 6: Model complexity on avMNIST. Best and second-best scores in **bold** and underlined.

Model	# Params (k)	Avg Acc (%) ↑	Speed (samples/s) ↑
mmFormer	204	96.56	1792
M³AE	206	95.89	1984
ShaSpec	195	96.11	**2112**
SimMLM	203	96.00	1120
LARGO (ours)	202	**97.00**	1984

Training. We employ a two-stage training protocol to ensure fair comparison across all models. In the first stage, modality-specific encoders (ImageEncoder and AudioEncoder) are pretrained independently for 100 epochs on their respective datasets: MNIST for visual features and FSDD for audio features. During the second stage, these pretrained encoders are frozen and integrated with model-specific fusion architectures. All models are then trained end-to-end for 100 epochs using random modality dropout.

For a fair comparison, the fusion bottleneck is 64 features for all models, the frozen encoders (approximately 74k parameters) are shared, the baseline-specific fusion architectures have approximately the same number of parameters, and all models use a batch size of 64, Adam with learning rate 0.001, and the cross-entropy loss.

[1] https://github.com/Jakobovski/free-spoken-digit-dataset

Quantitative results. Table 5 presents classification accuracy scores across all 3 missing-modality combinations. LARGO achieves the best average performance: +0.44% increase in accuracy compared to its closest competitors. This experiment shows that LARGO is flexible and can be extended to non-medical use cases, classification tasks, and heterogeneous modalities. Table 6 compares model complexity and efficiency: LARGO achieves the highest average accuracy (97%) at the same parameter budget.

5 Discussion of model complexity

Asymptotic complexity. For each convolutional layer, dynamic weight reconstruction performs a tensor contraction of cost $O(MC_{\text{in}}C_{\text{out}}KR)$, while the subsequent convolution costs $O(C_{\text{in}}C_{\text{out}}KS)$, where S is the spatial size of the feature map.

Stem and head layers. Stem and head are kept uncompressed. Their parameters grow exponentially with N, but on BraTS 2018 they account for 28,128 and 1,485 parameters respectively—only 0.131% of the total. The bulk of the budget lies in the inner convolutional layers, so the stem/head overhead remains modest as N grows. LARGO is therefore practical for $N \lesssim 10$, which covers most standard multimodal settings. For larger N, the stem can be padded to a fixed modality dimension and factorized along the model mode as well; we leave this to future work.

Table 7: Computational overhead of dynamic weight reconstruction in LARGO relative to a single nnU-Net backbone, measured on BraTS 2018 ($N = 4$, $M = 15$).

Metric	nnU-Net	LARGO (Ours)	Δ
Parameters	22,574,563	22,596,253	$+0.1\%$
GFLOPs	1,160.2	1,445.8	$+24.6\%$
Peak VRAM (MB)	4,910.79	6,144.98	$+25.1\%$

Empirical overhead. Table 7 compares LARGO to its underlying nnU-Net backbone on BraTS 2018. The total parameter count is essentially identical ($+0.1\%$): LARGO matches a single dedicated model in size while implicitly representing all $M = 15$ subset-specific models. Dynamic reconstruction increases inference compute by 24.6% in GFLOPs and peak training memory by 25.1%. This is acceptable since missing-modality models typically add operations (shared encoders, gating modules, etc.) on top of standard segmentation backbones. Inference cost varies by less than 1% with the number of available modalities (Table 11, supplementary material), since only the uncompressed stem depends on it.

Scaling with the number of modalities. Subset selection is an $O(1)$ row lookup in $\boldsymbol{A}$ (no learned router). With the rank recomputed from Equation (3), going from $M = 15$ to $M = 255$ subsets raises the parameter count by only 3.9% (stem/head share below 4%), while GFLOPs decrease (Table 12, supplementary material). The $1/M$ budget is a target rather than a constraint: even $R/4$ costs only about 1.6 Dice points (rank ablation, supplementary material). This is architectural evidence only; accuracy beyond four modalities remains to be validated.

6 Conclusion

We introduced LARGO, a principled hypernetwork framework for handling missing modalities that leverages CP decomposition to compress $2^N - 1$ dedicated models into a single network by exploiting structural similarities in convolutional kernels across modality-specific models.

Experimental evaluation on the BraTS 2018, ISLES 2022, and avMNIST datasets demonstrates that LARGO consistently outperforms state-of-the-art baselines across diverse missing-modality scenarios. In future work, we will explore more expressive, nonlinear hypernetwork formulations to further improve the weight approximation.

Limitations. Our empirical validation covers only convolutional backbones and $N \leq 4$ modalities. The CP factorization applies unchanged to matrix-valued weights (as in the linear layers of the avMNIST model), so transformer attention and MLP layers are covered algebraically but not validated; likewise, the scaling analysis in Section 5 does not establish that accuracy is preserved for $N > 4$. Dynamic reconstruction adds roughly 25% compute and memory, and the uncompressed stem and head still grow with $2^N - 1$ unless factorized.

Acknowledgments and Disclosure of Funding

This work is funded by the Flanders AI Research Program, under grant no. 174B09119. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government – department EWI. The authors declare no competing interests.

References

Pooya Ashtari, Frederik Maes, and Sabine Van Huffel. Low-rank convolutional networks for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 470–480. Springer, 2020.

Reza Azad, Mohammad Dehghanmanshadi, Nika Khosravi, Julien Cohen-Adad, and Dorit Merhof. Addressing missing modality challenges in MRI images: A comprehensive review. *Computational Visual Media*, 11(2):241–268, 2025. doi: 10.26599/CVM.2025.9450399.

Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification, 2021. URL `https://arxiv.org/abs/2107.02314`.

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4(1):170117, September 2017. ISSN 2052-4463. doi: 10.1038/sdata.2017.117. URL `https://www.nature.com/articles/sdata2017117`. Publisher: Nature Publishing Group.

Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2): 423–443, 2018.

Khaled Bayoudh, Raja Knani, Fayçal Hamdaoui, and Abdellatif Mtibaa. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *The Visual Computer*, 38(8):2939–2970, 2022.

Agisilaos Chartsias, Thomas Joyce, Rohan Dharmakumar, and Sotirios A Tsaftaris. Multimodal MR synthesis via modality-invariant latent representation. *IEEE transactions on medical imaging*, 37 (3):803–814, 2018.

Salman UH Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Çukur. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE transactions on medical imaging*, 38(10):2375–2388, 2019.

Ezequiel de la Rosa, Mauricio Reyes, Sook-Lei Liew, Alexandre Hutton, Roland Wiest, Johannes Kaesmacher, Uta Hanning, Arsany Hakim, Richard Zubal, Waldo Valenzuela, David Robben, Diana M. Sima, Vincenzo Anania, Arne Brys, James A. Meakin, Anne Mickan, Gabriel Broocks, Christian Heitkamp, Shengbo Gao, Kongming Liang, Ziji Zhang, Md Mahfuzur Rahman Siddiquee, Andriy Myronenko, Pooya Ashtari, Sabine Van Huffel, Hyunsu Jeong, Chiho Yoon, Chulhong Kim, Jiayu Huo, Sebastien Ourselin, Rachel Sparks, Albert Clèrigues, Arnau Oliver, Xavier Lladó, Liam Chalcroft, Ioannis Pappas, Jeroen Bertels, Ewout Heylen, Juliette Moreau, Nima Hatami, Carole Frindel, Abdul Qayyum, Moona Mazher, Domenec Puig, Shao-Chieh Lin, Chun-Jung Juan, Tianxi Hu, Lyndon Boone, Maged Goubran, Yi-Jui Liu, Susanne Wegener, Florian Kofler, Ivan Ezhov, Suprosanna Shit, Moritz R. Hernandez Petzsche, Michael Müller, Bjoern Menze, Jan S. Kirschke, and Benedikt Wiestler. DeepISLES: a clinically validated ischemic stroke segmentation model from the ISLES'22 challenge. *Nature Communications*, 16(1):7357, 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-62373-x.

Reuben Dorent, Samuel Joutard, Marc Modat, Sébastien Ourselin, and Tom Vercauteren. Hetero-modal variational encoder-decoder for joint modality completion and segmentation. In Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 74–82. Springer International Publishing, 2019. ISBN 978-3-030-32245-8. doi: 10.1007/978-3-030-32245-8_9.

David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.

Mohammad Havaei, Nicolas Guizard, Nicolas Chapados, and Yoshua Bengio. HeMIS: Hetero-modal image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 469–477. Springer, 2016. doi: 10.1007/978-3-319-46723-8_54.

Moritz R. Hernandez Petzsche, Ezequiel de la Rosa, Uta Hanning, Roland Wiest, Waldo Valenzuela, Mauricio Reyes, Maria Meyer, Sook-Lei Liew, Florian Kofler, Ivan Ezhov, David Robben, Alexandre Hutton, Tassilo Friedrich, Teresa Zarth, Johannes Bürkle, The Anh Baran, Björn Menze, Gabriel Broocks, Lukas Meyer, Claus Zimmer, Tobias Boeckh-Behrens, Maria Berndt, Benno Ikenberg, Benedikt Wiestler, and Jan S. Kirschke. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Scientific Data*, 9(1):762, 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01875-5. URL `https://www.nature.com/articles/s41597-022-01875-5`. Publisher: Nature Publishing Group.

Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021. ISSN 1548-7105. doi: 10.1038/s41592-020-01008-z. URL `https://www.nature.com/articles/s41592-020-01008-z`. Publisher: Nature Publishing Group.

Xu Jia, Bert De Brabandere, Tinne Tuytelaars, and Luc V Gool. Dynamic filter networks. *Advances in neural information processing systems*, 29, 2016.

Yong-Deok Kim, Eunhyeok Park, Sungjoo Yoo, Taelim Choi, Lu Yang, and Dongjun Shin. Compression of deep convolutional neural networks for fast and low power mobile applications. In *International Conference on Learning Representations*, 2015.

Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500, 2009. doi: 10.1137/07070111X. URL `https://doi.org/10.1137/07070111X`.

V Lebedev, Y Ganin, M Rakhuba, I Oseledets, and V Lempitsky. Speeding-up convolutional neural networks using fine-tuned CP-decomposition. In *3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings*, 2015.

Sijie Li, Chen Chen, and Jungong Han. SimMLM: A simple framework for multi-modal learning with missing modality. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 24068–24077, 2025.

Hong Liu, Dong Wei, Donghuan Lu, Jinghan Sun, Liansheng Wang, and Yefeng Zheng. M3AE: Multimodal representation learning for brain tumor segmentation with missing modalities. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 1657–1665, 2023.

Zhun Liu, Ying Shen, Varun Bharadhwaj Lakshminarasimhan, Paul Pu Liang, AmirAli Bagher Zadeh, and Louis-Philippe Morency. Efficient low-rank multimodal fusion with modality-specific factors. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2247–2256, 2018.

Mengmeng Ma, Jian Ren, Long Zhao, Sergey Tulyakov, Cathy Wu, and Xi Peng. SMIL: Multimodal learning with severely missing modality. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 2302–2310, 2021.

Harsh Maheshwari, Yen-Cheng Liu, and Zsolt Kira. Missing modality robustness in semi-supervised multi-modal semantic segmentation. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1009–1019, 2024. doi: 10.1109/WACV57701.2024.00106.

Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, et al. The multimodal brain tumor image segmentation benchmark (BraTS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015. doi: 10.1109/TMI.2014.2377694.

Elvis Nava, Seijin Kobayashi, Yifei Yin, Robert K Katzschmann, and Benjamin F Grewe. Meta-learning via classifier (-free) diffusion guidance. *arXiv preprint arXiv:2210.08942*, 2022.

Hugh G. Pemberton, Jiaming Wu, Ivar Kommers, Domenique M. J. Müller, Yipeng Hu, Olivia Goodkin, Sjoerd B. Vos, Sotirios Bisdas, Pierre A. Robe, Hilko Ardon, Lorenzo Bello, Marco Rossi, Tommaso Sciortino, Marco Conti Nibali, Mitchel S. Berger, Shawn L. Hervey-Jumper, Wim Bouwknegt, Wimar A. Van den Brink, Julia Furtner, Seunggu J. Han, Albert J. S. Idema, Barbara Kiesel, Georg Widhalm, Alfred Kloet, Michiel Wagemakers, Aeilko H. Zwinderman, Sandro M. Krieg, Emmanuel Mandonnet, Ferran Prados, Philip de Witt Hamer, Frederik Barkhof, and Roelant S. Eijgelaar. Multi-class glioma segmentation on real-world data with missing MRI sequences: comparison of three deep learning algorithms. *Scientific Reports*, 13(1):18911, November 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-44794-0. URL `https://www.nature.com/articles/s41598-023-44794-0`. Publisher: Nature Publishing Group.

Junjie Shi, Zhaobin Sun, Li Yu, Xin Yang, and Zengqiang Yan. Addressing imbalanced modal incompleteness in realistic multi-modal medical image segmentation via hierarchical gradient alignment. *IEEE Transactions on Medical Imaging*, 45(6):3066–3081, 2026. doi: 10.1109/TMI.2026.3667605.

Nicholas D. Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E. Papalexakis, and Christos Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017. ISSN 1053-587X, 1941-0476. doi: 10.1109/TSP.2017.2690524. URL `http://arxiv.org/abs/1607.01668`.

Peibo Song, Xiaotian Xue, Jinshuo Zhang, Zihao Wang, Jinhua Liu, Shujun Fu, Fangxun Bao, and Si Yong Yeo. Uni-encoder meets multi-encoders: Representation before fusion for brain tumor segmentation with missing modalities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Hu Wang, Yuanhong Chen, Congbo Ma, Jodie Avery, Louise Hull, and Gustavo Carneiro. Multi-modal learning with missing modality via shared-specific feature modelling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15878–15887, 2023a. doi: 10.1109/CVPR52729.2023.01524.

Hu Wang, Congbo Ma, Jianpeng Zhang, Yuan Zhang, Jodie Avery, Louise Hull, and Gustavo Carneiro. Learnable Cross-modal Knowledge Distillation for Multi-modal Learning with Missing Modality. In Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, and Russell Taylor, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 216–226, Cham, 2023b. Springer Nature Switzerland. ISBN 978-3-031-43901-8. doi: 10.1007/978-3-031-43901-8_21.

Renjie Wu, Hu Wang, Hsiang-Ting Chen, and Gustavo Carneiro. Deep multimodal learning with missing modality: A survey, 2024. URL `https://arxiv.org/abs/2409.07825`.

Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. Tensor fusion network for multimodal sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 1103–1114, 2017.

Chris Zhang, Mengye Ren, and Raquel Urtasun. Graph hypernetworks for neural architecture search. *arXiv preprint arXiv:1810.05749*, 2018.

Yao Zhang, Nanjun He, Jiawei Yang, Yuexiang Li, Dong Wei, Yawen Huang, Yang Zhang, Zhiqiang He, and Yefeng Zheng. mmFormer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 107–117. Springer, 2022.

Yi Zhang, Guoxia Xu, Meng Zhao, Hao Wang, Fan Shi, and Shengyong Chen. Tdsf-net: Tensor decomposition-based subspace fusion network for multimodal medical image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2025a.

Zheyu Zhang, Yayuan Lu, Feipeng Ma, Yueyi Zhang, Huanjing Yue, and Xiaoyan Sun. Incomplete multi-modal brain tumor segmentation via learnable sorting state space model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025b. doi: 10.1109/CVPR52734.2025.02420.

A Backbone architecture

This section describes the base network architectures used in the BraTS 2018, ISLES 2022, and avMNIST experiments. Both architectures employ low-rank tensor decomposition in their convolutional and linear layers to enable parameter-efficient handling of missing modalities.

A.1 3D segmentation network architecture

For the BraTS 2018 and ISLES 2022 datasets, we employ a U-Net architecture based on the nnU-Net framework [Isensee et al., 2021]. The network follows the standard nnU-Net configuration with a symmetric encoder-decoder structure.

BraTS 2018 dataset. The encoder consists of four downsampling stages with progressively increasing channel dimensions (32, 64, 128, 256, 512) and decreasing spatial resolutions ($128^3 \rightarrow 64^3 \rightarrow 32^3 \rightarrow 16^3 \rightarrow 8^3$). Each encoder block contains two 3D convolutions ($3 \times 3 \times 3$ kernels) with instance normalization and leaky ReLU activations, followed by strided convolution for downsampling.

The decoder mirrors the encoder structure with four upsampling stages using transposed convolutions. The final output layer produces a single-channel segmentation mask at the original 128^3 resolution.

The network architecture is visually represented in Figure 4.

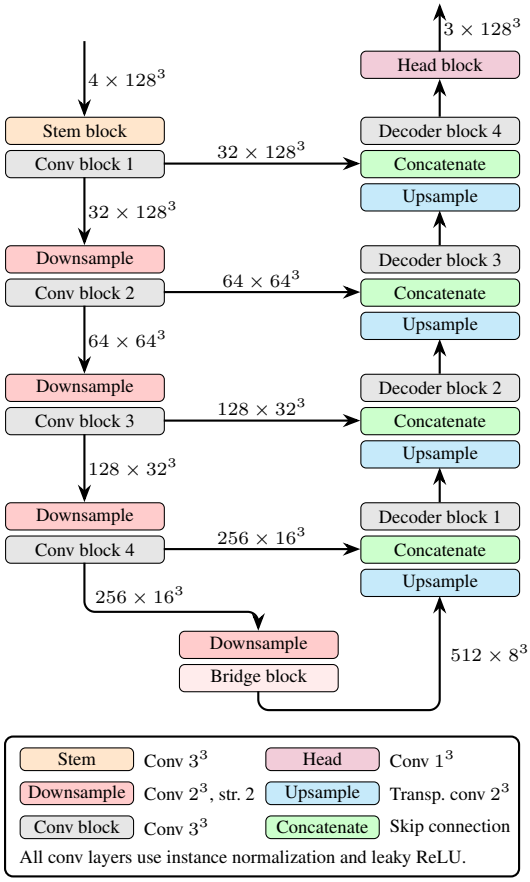


Figure 4: Configuration of the nnU-Net architecture used for the BraTS 2018 dataset. The encoder path (left) progressively downsamples the input through conv blocks with increasing channels. Skip connections transfer encoder features to the decoder path (right), which upsamples back to the original resolution. Numbers indicate channels×spatial resolution.

ISLES 2022 dataset. For the ISLES 2022 dataset, the network has one fewer downsample and upsample layer. The downsampling layers progressively increase the channel dimensions (64, 128, 256, 512) and decrease the spatial resolution ($64^3 \rightarrow 32^3 \rightarrow 16^3 \rightarrow 8^3$).

A.2 avMNIST network architecture

For the avMNIST digit classification task, we employ a fusion-based architecture that processes two modalities (image and audio) and handles $2^2 - 1 = 3$ modality combinations (image-only, audio-only, and both).

The architecture follows a modality-specific encoding followed by fusion paradigm, illustrated in Figure 5.

Modality-specific encoders. Separate convolutional encoders process each modality independently. The image encoder processes 28×28 grayscale MNIST images through a 4-layer CNN with progressively increasing channels (16, 32, 64, 128), producing a 160-dimensional feature vector. The audio encoder processes MFCC spectrograms through a similar 4-layer CNN architecture with adapted dimensions, producing a 320-dimensional feature vector.

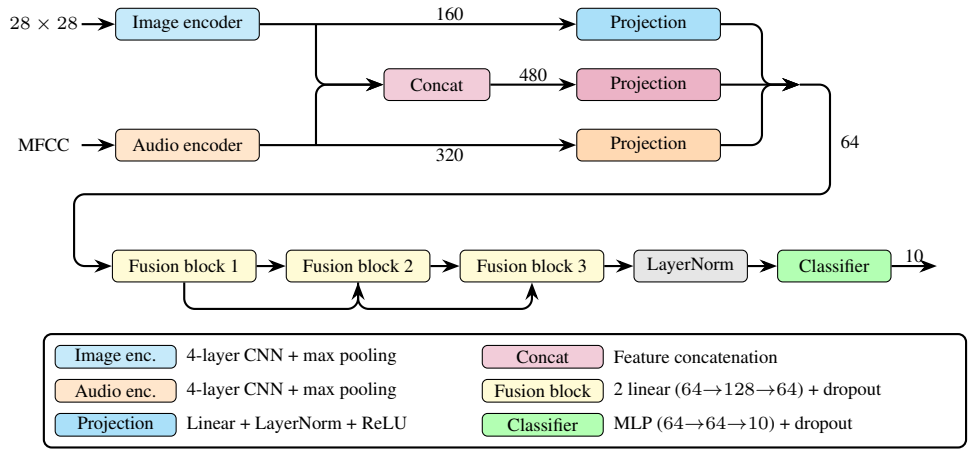


Figure 5: Configuration of the fusion architecture used for the avMNIST dataset. The image and audio encoders independently extract features. Features from available modalities are concatenated and projected on a 64-dimensional feature vector. Three fusion blocks with residual connections process the combined features, followed by layer normalization and a classifier producing digit predictions.

Projection and fusion. The inputs from the available modalities are concatenated and projected by a unique projection block on a common 64-dimensional feature vector.

Deep fusion network. The projected features pass through three fusion blocks with residual connections. Each block consists of two linear transformations ($64 \rightarrow 128 \rightarrow 64$) with layer normalization, ReLU activations, and dropout for regularization. A final layer normalization is applied before classification.

Classification head. The classification head produces final predictions for the different digit classes. Each head consists of a two-layer MLP with dropout, allowing the network to learn combination-specific decision boundaries.

CPD compression of linear layers. For linear layers, the joint weight tensor can be represented as $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{M \times C_{\text{in}} \times C_{\text{out}}}$. The CPD decomposition is then similar to Equation (2), discarding the matrix $\boldsymbol{D}$. The rank selection is equivalent to Equation (3), discarding K.

B Ablation studies

This section analyzes the impact of compression-related hyperparameters on model performance. We examine rank selection and compare different tensor decomposition methods on the ISLES 2022 dataset using 5-fold cross-validation.

B.1 Impact of rank on performance

Table 8: Dice scores (%) ↑ on ISLES 2022 (5-fold cross-validation) for different values of the rank R and for the dedicated models (Ded.). •/○: modality present/absent.

DWI	ADC	FLAIR	Stroke Lesion					
			$R/4$	$R/2$	R	$2R$	$7R$	Ded.
•	○	○	73.52	74.79	74.23	74.96	74.79	74.22
○	•	○	39.74	41.12	42.49	43.06	43.95	45.08
○	○	•	38.35	39.27	41.09	41.91	41.73	41.35
•	•	○	74.58	75.23	74.67	75.32	76.10	75.11
•	○	•	73.59	74.57	75.75	75.49	75.63	74.73
○	•	•	51.80	52.80	53.70	53.92	54.83	54.90
•	•	•	75.92	76.85	76.82	77.14	76.45	74.89
Average			61.07	62.09	62.68	63.11	63.36	62.90

The rank R directly controls the compression ratio in our decomposition. As defined in Section 3, we set R to achieve a compression ratio of $1/M$ relative to dedicated models, where M is the number of modality combinations. Lower ranks reduce parameters but limit model capacity, while higher ranks increase parameters.

Figure 6 and Table 8 present results for five rank settings: $R/4$, $R/2$, R (our default), $2R$, and $7R$. We also include dedicated models for comparison, which have approximately the same parameter

count as the $7R$ configuration. The dedicated models were trained as a hypernetwork, with separate models and the same training procedure. The average Dice score increases from 61.07% at $R/4$ to 63.36% at $7R$, while the Hausdorff^{95} distance decreases from 11.51mm to 10.0mm. This trend indicates that higher compression (lower rank) reduces accuracy, as expected.

At $7R$, LARGO achieves 63.36% Dice score compared to 62.90% for dedicated models with similar parameter counts. The default rank R achieves 62.68% Dice score with $1/M$ compression relative to dedicated models.

B.2 Comparison of tensor decompositions

We compare the canonical polyadic decomposition (CPD) used in our method against the Tucker decomposition, another common tensor decomposition approach. The Tucker decomposition, illustrated in Figure 7, uses a core tensor and factor matrices along each mode. The modes M and K are not compressed, since these are relatively small compared to C_{in} and C_{out}. The modes C_{in} and C_{out} share the same rank R. We set the rank for the $1/M$ compression as follows:

$$R = \left\lfloor \frac{-(C_{\text{in}} + C_{\text{out}}) + \sqrt{\Delta}}{2MK} \right\rfloor, \qquad (6)$$

with discriminant $\Delta = (C_{\text{in}} + C_{\text{out}})^2 - 4MK\left(M^2 + K^2 - K\, C_{\text{in}} C_{\text{out}}\right) \geq 0$.

Table 9 compares CPD and Tucker decompositions with matched parameter budgets. CPD achieves an average Dice score of 62.68% compared to 61.97% for Tucker. The difference is larger for single-modality scenarios: FLAIR-only shows 41.09% vs 37.80% (3.29 percentage points), and ADC+FLAIR shows 53.70% vs 52.35% (1.35 points). For the complete three-modality case, both methods perform similarly (76.82% vs 77.07%).

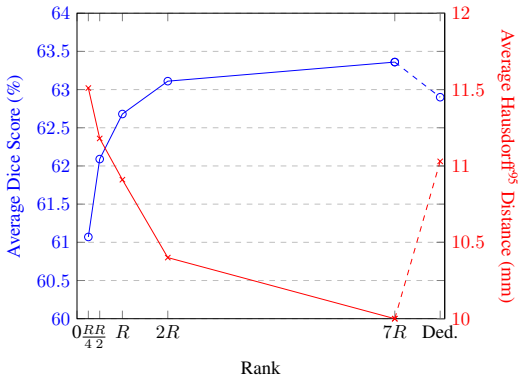


Figure 6: Ablation study on the rank R for the ISLES 2022 dataset using 5-fold cross-validation. We present the values $R/4$, $R/2$, R, $2R$, and $7R$, with R as in Equation (3), as well as the performance of the dedicated models (Ded.), which is approximately equivalent in terms of number of parameters to the $7R$ case. The points for the average Dice score (↑) and Hausdorff^{95} distance (↓) are represented in blue and red, respectively.

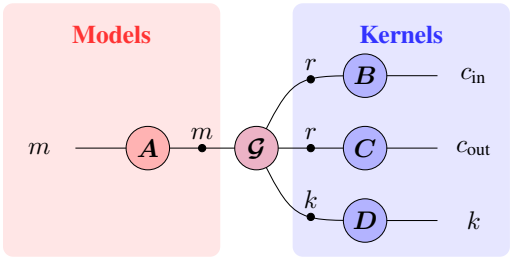


Figure 7: Tensor diagram of the Tucker reparameterization of the convolutional and transposed convolutional layers.

Table 9: Comparison of Dice scores (%) ↑ on the ISLES 2022 dataset using 5-fold cross-validation for CPD and Tucker decomposition. Symbols • and ○ denote the presence and absence of a modality, respectively.

DWI	ADC	FLAIR	Stroke Lesion	
			CPD	Tucker
•	○	○	74.23	74.08
○	•	○	42.49	42.65
○	○	•	41.09	37.80
•	•	○	74.67	74.62
•	○	•	75.75	75.19
○	•	•	53.70	52.35
•	•	•	76.82	77.07
	Average		62.68	61.97

C Statistical analysis

Table 10 reports the patient-level comparison of LARGO against each baseline described in Section 4. For every patient, the Dice score is averaged over all modality subsets (and tumor regions for BraTS 2018), which yields one paired observation per patient and method. We apply two-sided paired Wilcoxon signed-rank tests and correct the p-values with the Benjamini–Hochberg procedure. All aggregate comparisons remain significant after correction; at the level of individual modality subsets, 57 of 60 comparisons on BraTS 2018 and 26 of 28 on ISLES 2022 remain significant.

Table 10: Patient-level statistical comparison of LARGO against each baseline. ΔDice is the average gain of LARGO in percentage points; "won" is the number of patients for which LARGO obtains the higher subset-averaged Dice score; p-values are from two-sided paired Wilcoxon signed-rank tests after Benjamini–Hochberg correction.

LARGO vs.	**BraTS 2018** (285 patients)			**ISLES 2022** (250 patients)		
	ΔDice	Won	Corrected p	ΔDice	Won	Corrected p
mmFormer	$+3.13$	243/285	2.7×10^{-29}	$+5.79$	198/250	1.3×10^{-21}
M³AE	$+2.05$	223/285	2.0×10^{-19}	$+2.53$	181/250	3.1×10^{-14}
ShaSpec	$+0.69$	185/285	5.8×10^{-6}	$+11.75$	240/250	3.6×10^{-40}
SimMLM	$+1.82$	207/285	9.2×10^{-16}	$+2.65$	190/250	1.9×10^{-15}

D Inference cost and architectural scaling

This section complements Section 5 of the main paper.

Table 11 reports GFLOPs and inference time per 128^3 patch of LARGO on BraTS 2018 as a function of the number of available modalities. Both vary by less than 1% (1144.3–1155.1 GFLOPs, 65.8–66.6 ms), since only the uncompressed stem depends on the number of input channels; reconstructing the inner-layer kernels and applying them cost the same for every subset. The four-modality value is the one reported in Table 2.

Table 11: Inference cost of LARGO vs. number of available modalities (BraTS 2018, 128^3 patch).

Available modalities	GFLOPs	Time per patch
1	1144.3	65.8 ms
2	1147.9	66.1 ms
3	1151.5	66.2 ms
4	1155.1	66.6 ms

Table 12 reports how parameters, the stem/head share, GFLOPs, and the per-layer rank range change when the number of modalities N grows and the rank is recomputed from Equation (3). Going from $N = 4$ to $N = 8$ ($M = 255$ subsets) raises the parameter count by only 3.9%, while GFLOPs decrease with the lower ranks. At $N = 10$ the uncompressed stem and head reach 16.8% of the budget, which motivates factorizing them as discussed in Section 5. These numbers are architectural evidence and do not establish that accuracy is preserved for $N > 4$.

Table 12: Architectural scaling of LARGO with the number of modalities N (BraTS 2018 backbone), with the per-layer rank recomputed from Equation (3).

N	$M = 2^N - 1$	Params (M)	Stem/head share	GFLOPs	Rank range
4	15	22.60	0.13%	1155.1	137–6639
6	63	22.74	0.77%	1151.5	98–6353
8	255	23.48	3.91%	1126.1	45–5419
10	1023	27.11	16.81%	1072.3	14–3412

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state that LARGO compresses the family of $2^N - 1$ dedicated models into a single low-rank hypernetwork via CP decomposition. These claims are supported by the method (Section 3), the complexity analysis (Section 5), and the experiments on BraTS 2018, ISLES 2022, and avMNIST.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: A dedicated Limitations paragraph at the end of the paper discusses the restriction to convolutional backbones and $N \leq 4$ modalities, the architectural (not empirical) nature of the scaling analysis, the overhead of dynamic weight reconstruction, and the simplicity of avMNIST. Section 5 discusses scalability limits (LARGO is practical for $N \lesssim 10$ due to the uncompressed stem/head growing linearly with $M = 2^N - 1$) and reports the $\sim 25\%$ compute and memory overhead of dynamic weight reconstruction. Evaluation is concentrated on convolutional backbones and medical imaging, with avMNIST included only as a proof-of-concept for non-medical settings.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: The paper does not include formal theorems or proofs; the contribution is methodological and empirical.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The hypernetwork construction and training procedure are described in Section 3; datasets, training schedules, optimizers, learning rates, and weight decay are reported per dataset in the Experiments section; architecture diagrams and the fusion network for avMNIST are provided in the supplementary material.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: All datasets used (BraTS 2018, ISLES 2022, avMNIST) are publicly available. The source code, including training and evaluation scripts and instructions to reproduce the main results, is openly available at `https://github.com/nielsvyncke/LARGO`; the supplementary material additionally provides architectural details and hyperparameters.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The Experiments section reports data splits (5-fold cross-validation on BraTS and ISLES, predefined train/test split on avMNIST), optimizer (SGD with Nesterov momentum 0.99 for medical, Adam for avMNIST), learning rate, weight decay, batch size, and number of epochs. Architectural hyperparameters (channels, depth) are specified in the supplementary material.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Results on BraTS 2018 and ISLES 2022 are reported as mean $\pm$ standard deviation across 5-fold cross-validation in Tables 1 to 4. In addition, Table 10 reports patient-level two-sided paired Wilcoxon signed-rank tests of LARGO against each baseline with Benjamini–Hochberg correction for multiple comparisons.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Section 5 reports parameter count, GFLOPs, peak training VRAM, and per-epoch training time for LARGO and the baselines on BraTS 2018, together with inference speed in Tables 2 and 4.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

Justification: The research uses publicly available, de-identified datasets (BraTS, ISLES, avMNIST). It conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: Positive: improving robustness to missing modalities can make multimodal medical imaging tools more practical in clinical settings where some MRI sequences may be unavailable. Negative: like any medical-imaging model, LARGO should not be used for clinical decision-making without validation and regulatory approval, and predictions on heavily missing modalities should be interpreted with care.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release pre-trained generative models, scraped datasets, or other artifacts with a high risk of misuse.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All datasets (BraTS 2018, ISLES 2022, avMNIST) and the nnU-Net framework are cited to their original publications, used under their public licenses, and respected in their terms of use.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new datasets or pre-trained models. The released source code (`https://github.com/nielsvyncke/LARGO`) is documented with usage instructions.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research with human subjects; it uses publicly available, previously collected datasets.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve new data collection from human subjects; it uses publicly available datasets that were collected and released under their own ethical approvals.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not used as a component of the core methodology; any use was limited to writing and editing assistance.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.