Towards Data-Efficient Medical Imaging: A Generative and Semi-Supervised Framework

Mosong Ma[1] Tania Stathaki[1] Michalis Lazarou[2]

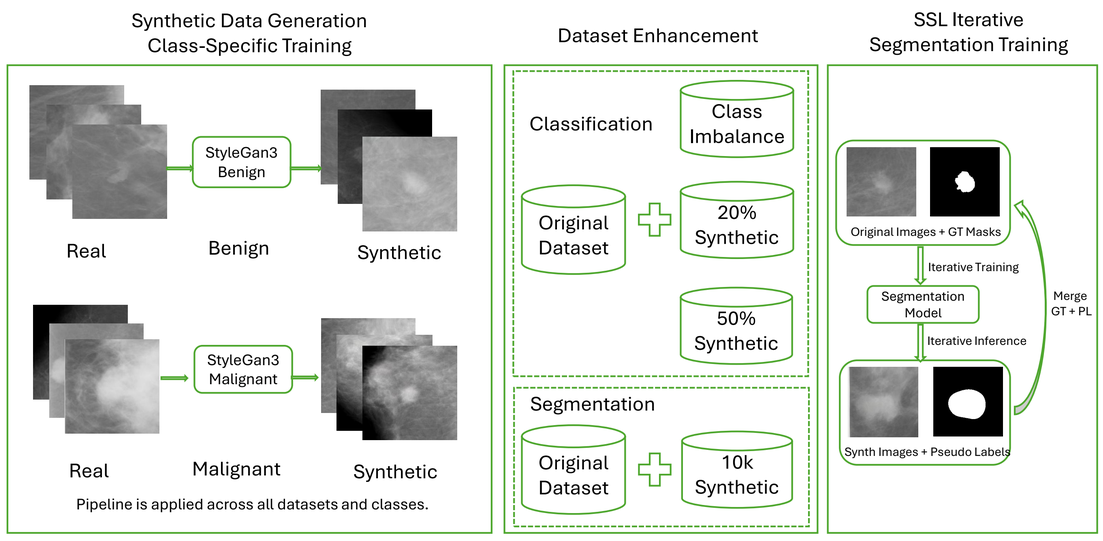
[1]Imperial College London [2]Centre for Vision, Speech and Signal Processing, University of Surrey

Motivation

- Deep learning in medical imaging is constrained by scarce labels and class imbalance.
- Synthetic data and semi-supervised learning (SSL) can expand training sets without extra manual annotation.
- **Goal:** Improve classification and segmentation under limited supervision with a practical, reproducible recipe.

SSGNet: Overview

- Train **class-specific StyleGAN3** generators to synthesize high-fidelity images per class.
- For classification, the class-specific generation can enlarge original dataset with synthetic data and their corresponding class label.
- For segmentation, obtain **pseudo-masks** for synthetic images via a baseline model and **iterative refinement**.
- Plug into common backbones (*e.g.*, ResNet-50, VM-UNet, Adaptive t-vMF Dice) as a **drop-in data enhancement**.



Contributions

- A unified framework combining **class-specific** generation and **iterative SSL** for segmentation.
- Demonstrate consistent gains across CBIS-DDSM, Kvasir-SEG, ISIC 2017/2018, Chest X-ray, BreastMNIST.
- Addresses class imbalance and small sample size by training per-class generators that equalize rare classes and stabilize downstream training.

Datasets

Task	Datasets
Classification	Chest X-ray, BreastMNIST, CBIS-DDSM, Kvasir-SEG
Segmentation	CBIS-DDSM, Kvasir-SEG, ISIC 2017/2018

Ablation / Analysis

- **FID** suggests best synthesis for Chest X-ray (27–29), dermoscopy (35–40); mammography is harder (48–53).
- **Iterations:** 1–2 SSL rounds capture most of the gains; diminishing returns beyond.
- **Backbone-agnostic:** Gains hold for ResNet-50, VM-UNet and Adaptive t-vMF Dice Loss.

Takeaways

- **Practical:** No change to model architectures; treat SSGNet as a data layer.
- **Robust:** Helps both classification and segmentation under limited labels.
- **Reproducible:** Works across multiple public datasets.

Method Details

1) Class-specific StyleGAN3. Train one generator per class to avoid imbalance artifacts; synthesize diverse, high-quality samples.

2) Baseline models. ResNet-50 for classification; VM-UNet and Adaptive t-vMF Dice for segmentation.

3) Iterative pseudo-labeling (segm.).

1. Train baseline on real labeled data.
2. Generate synthetic images; predict masks (*pseudo-labels*).
3. Retrain on real + pseudo; re-predict masks to replace and repeat for t rounds (we use t=2).

Contact

mosong.ma18@imperial.ac.uk t.stathaki@imperial.ac.uk michalislazarou93@gmail.com

Key Results — Segmentation

Dataset	Model	mIoU	Dice	Acc.	Spe.	Sen.
ISIC2017	VM-UNet (baseline)	78.80	88.10	96.10	97.40	88.10
	+SSGNet (ours)	**80.54**	**89.23**	**96.44**	**98.10**	**88.15**
	AtvMF (baseline)	84.63	91.43	94.59	96.31	87.40
	+SSGNet (ours)	**86.90**	**92.81**	**95.45**	**96.76**	**89.98**
ISIC2018	VM-UNet (baseline)	80.22	89.04	94.81	96.10	90.33
	+SSGNet (ours)	**81.54**	**89.83**	**95.02**	**96.70**	**91.80**
	AtvMF (baseline)	84.44	91.42	94.02	96.59	85.32
	+SSGNet (ours)	**87.10**	**92.96**	**95.11**	**97.45**	**87.20**
CBIS-DDSM	VM-UNet (baseline)	52.51	68.86	94.99	95.60	64.22
	+SSGNet (ours)	**56.99**	**72.60**	**95.25**	**97.20**	**74.15**
	AtvMF (baseline)	59.63	71.10	87.20	91.28	55.33
	+SSGNet (ours)	**60.98**	**72.22**	**88.51**	**93.10**	**74.15**
Kvasir-SEG	VM-UNet (baseline)	74.51	85.37	95.50	96.22	84.14
	+SSGNet (ours)	**76.94**	**87.02**	**95.77**	**97.53**	**87.03**
	AtvMF (baseline)	86.35	92.37	95.99	98.70	82.40
	+SSGNet (ours)	**86.79**	**92.70**	**96.15**	**98.87**	**82.50**

Table: Comparison of segmentation results across datasets. Bold indicates improved scores from SSGNet over each baseline.

Key Results — Classification

Dataset	Model	Accuracy (%)	Macro F1 (%)
Kvasir-SEG	ResNet50 (baseline)	86.0	86.5
	+SSGNet (ours)	**91.0**	**91.0**
Chest X-ray	ResNet50 (baseline)	77.0	70.5
	+SSGNet (ours)	**84.0**	**80.5**
BreastMNIST	ResNet50 (baseline)	81.0	70.0
	+SSGNet (ours)	**87.0**	**82.5**
CBIS-DDSM	ResNet50 (baseline)	60.0	56.5
	+SSGNet (ours)	**63.0**	**61.5**

Table: Classification performance across datasets. Bold indicates improvement from SSGNet over each baseline.

Synthetic augmentation helps most where classes are imbalanced or scarce.

Segmentation Progression (All Stages Shown Per Row)

Dataset	Model	mIoU (%)	Dice (%)	Accuracy (%)
ISIC2017	VM-UNet	78.8 → 79.9 → **80.5**	88.1 → 88.9 → **89.2**	96.1 → 96.3 → **96.4**
ISIC2017	AtvMF	84.6 → **86.9** → 85.3	91.4 → **92.8** → 91.8	94.6 → **95.5** → 94.8
ISIC2018	VM-UNet	80.2 → 81.1 → **81.5**	89.0 → 89.6 → **89.8**	94.8 → 94.9 → **95.0**
ISIC2018	AtvMF	84.6 → 86.4 → **87.1**	91.4 → 92.5 → **93.0**	94.0 → 94.8 → **95.1**
Kvasir-SEG	VM-UNet	74.5 → 76.1 → **76.9**	85.4 → 86.4 → **87.0**	95.5 → 95.8 → **95.8**
Kvasir-SEG	AtvMF	86.4 → **86.8** → 82.4	92.4 → **92.7** → 92.5	96.0 → **96.2** → 94.6
CBIS-DDSM	VM-UNet	52.5 → 56.7 → **57.0**	68.9 → 72.4 → **72.6**	95.0 → 95.4 → **95.4**
CBIS-DDSM	AtvMF	59.6 → 60.5 → **61.0**	71.1 → 71.8 → **72.2**	87.2 → 88.0 → **88.5**

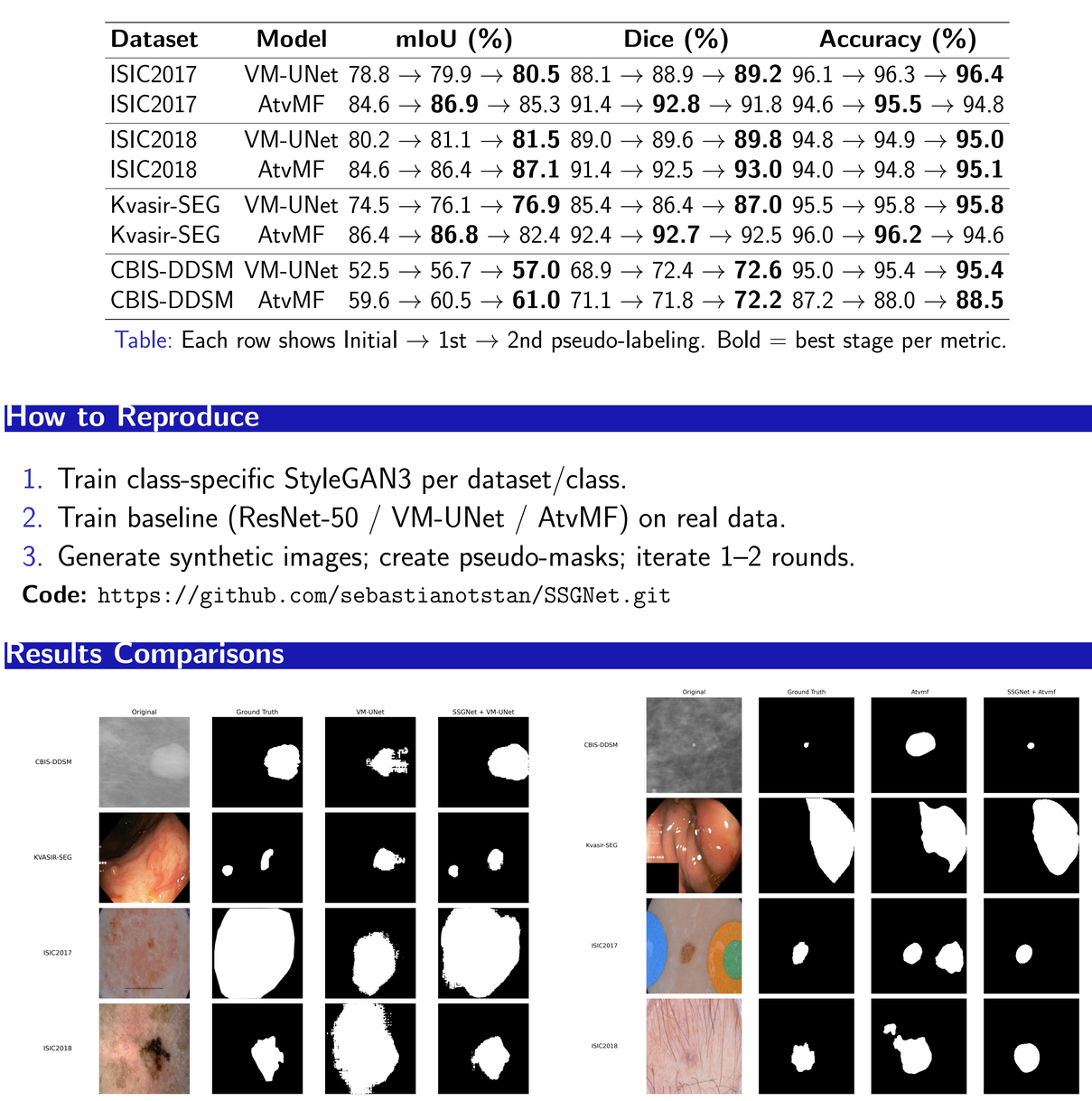
Table: Each row shows Initial → 1st → 2nd pseudo-labeling. Bold = best stage per metric.

How to Reproduce

1. Train class-specific StyleGAN3 per dataset/class.
2. Train baseline (ResNet-50 / VM-UNet / AtvMF) on real data.
3. Generate synthetic images; create pseudo-masks; iterate 1–2 rounds.

Code: `https://github.com/sebastianotstan/SSGNet.git`

Results Comparisons



*(a) Comparison with VM-UNet

*(b) Comparison with Adaptive t-vMF Dice Loss

Future Work

- Extend to **multi-modal data** (MRI + PET) to study disease progression.
- Integrate diffusion-based generation for fine-grained tissue realism.
- Explore automatic confidence-weighted pseudo-label selection.
- Benchmark under varying label budgets for scalability analysis.