

LiteAugNet: A Lightweight Semantic-Guided Augmentation Network for Efficient Edge-Level Image Classification

Mohammad Shahedur Rahman



Mohammad Tahmid Bari



Md. Nasim Adnan



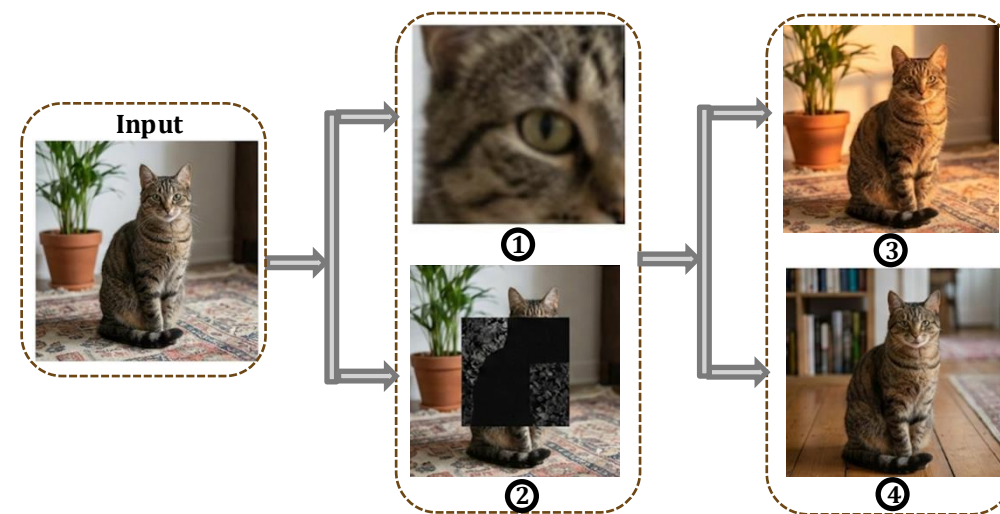
**JASHORE UNIVERSITY OF
SCIENCE AND TECHNOLOGY**

Arshad Parvez



Data Augmentation

- Data augmentation refers to
 - a **technique** that artificially expand the training data
 - applying **label-preserving transformation** to improve **generalization** and **robustness**.
- **Geometric**: flip, crop, rotation, scaling
- **Photometric**: color jitter, noise, blur.
- **Mixing-based**: Mixup, CutMix, SaliencyMix.
- **Generative**: GAN-based, diffusion-based.



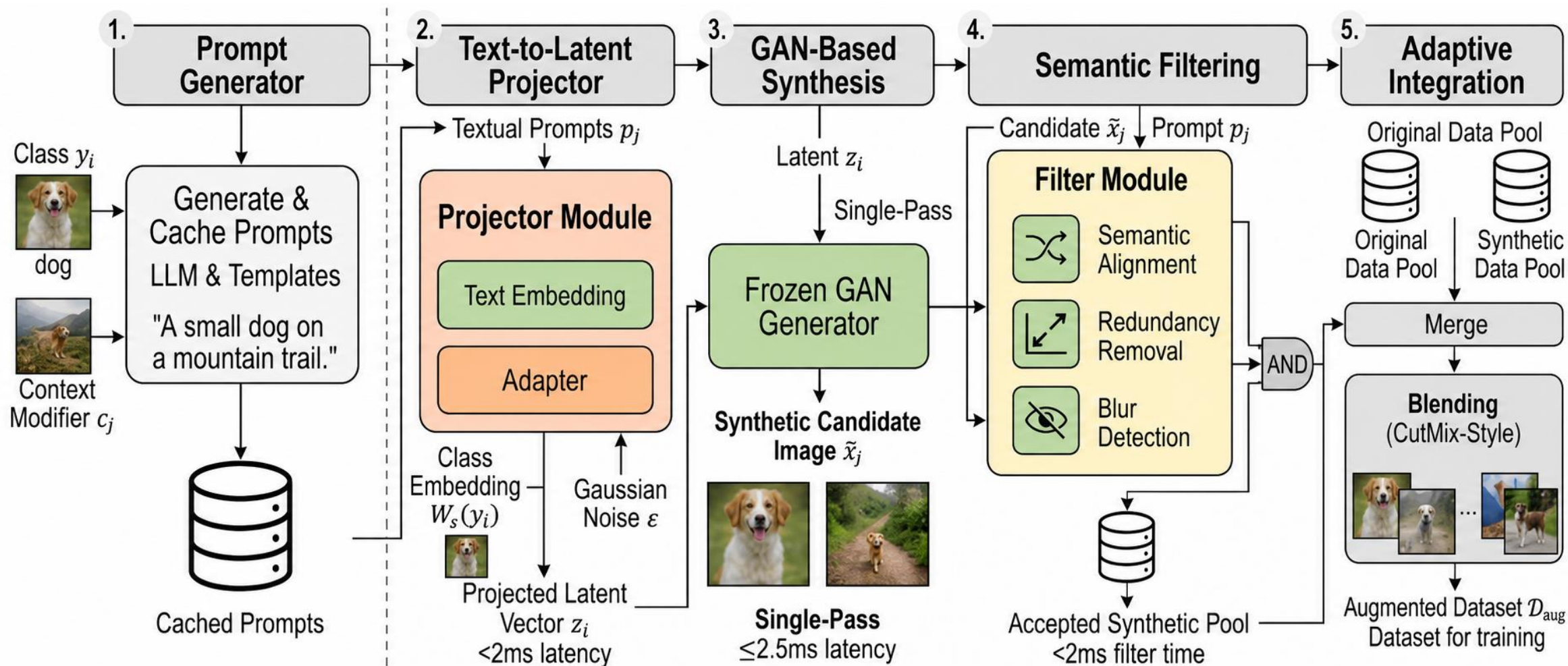
Why Semantic Augmentation Matters for Edge AI?

- Motivation:
 - Diffusion models require *expensive iterative sampling*.
 - Large Vision Language Models (VLMs) exceed *edge memory constraint*.
 - Most of the generators target *256 x 256 instead of 32 x 32*.
 - *No semantic filtering* leads to off-class images, duplicates, blurry samples.
 - *An effective study on semantic augmentation is required for Edge AI.*

Our Contribution: LiteAugNET

- LiteAugNet: a lightweight **five-stage** semantic-guided augmentation framework designed specifically for edge devices .
- Lightweight semantic synthesis
 - *First framework to integrate* compact vision-language encoder and a quantized LoRA adapter to *steer a frozen StyleGAN2-ADA generator*.
- Efficient on-device semantic filtering
 - *Introduce a three-step filtering pipeline* combining *semantic alignment, redundancy removal* and *blur detection* to retain high-quality synthetic samples.
- Robust and edge-efficient learning
 - *Adaptively integrates* filtered synthetic data with real data to improve *accuracy, robustness and real-time edge deployment*.

Overview of LiteAugNET



LiteAugNET: Stage I - Semantic Prompt Generation

- Class-aware prompt generation
 - Generate $k = 10$ *descriptive prompt per class* using **Phi-3-mini-4k-instruct (INT8)**.
- Contextual semantic expansion
 - Prompts incorporate environment cues like *lighting*, *weather*, and *background variations* to enrich semantic diversity.
- Offline prompt generation
 - All prompts are generated *offline* to eliminate on-device inference latency.

LiteAugNET: Stage 2-Text-to-Latent Projection

- Compact vision-language encoding
 - Encode prompts using *MobileCLIP* or *SIGLIP-Tiny* to obtain semantic text embeddings.
- Quantized LoRA-style projection
 - Map text embeddings into the latent space using *quantized LoRA-style adapter* ($\leq 50K$ *INT8 parameters*).
- Class-aware latent conditioning
 - Combine projected latent vectors with *learned class embeddings* for controllable semantic generation.

LiteAugNET: Stage 3-GAN-Based Semantic Synthesis

- Single-pass semantic generation
 - Generate 32×32 semantic images using a *frozen StyleGAN2-ADA* generator.
- Efficient edge deployment
 - Optimize the generator with *TensorRT* to achieve ≤ 2.5 ms synthesis latency per sample.
- Label-consistent synthesis
 - Produce synthetic images that preserve the *semantic intent* of the textual prompts.

LiteAugNET: Stage 4- On-Device Semantic Filtering

- Semantic alignment
 - Retain samples satisfying *CLIP cosine similarity* above a semantic threshold.
- Redundancy removal
 - Discard **near-duplicate** images using *SSIM-based filtering*.
- Visual quality assessment
 - Reject blurry or low-quality images using *Laplacian variance*.
- High-quality synthetic pool
 - Only samples *passing all filtering stages* are added to the augmented dataset.

LiteAugNET: Stage 5-Adaptive Data Integration

- Real-synthetic data fusion
 - Merge validated synthetic images with the original dataset using *a fixed ratio (= 0.3)*.
- Mixing-based augmentation
 - Apply *Cutmix* and *SalientMix* to combine semantic richness with visual fidelity.
- Robust model training
 - Improve feature diversity, generalization and robustness through adaptive data integration.

LiteAugNET: Experimental Setup

- Dataset and Models

- CIFAR 10, Tiny-Imagenet, SVHN
evaluated at 32 x 32 RGB with ResNet18,
MobileNetv2, VGG-16 and ViT-Tiny.

Dataset	Cls	Train	Test	N _{syn}
CIFAR-10	10	50,000	10,000	2,142
Tiny-ImageNet	200	100,000	10,000	214
SVHN	10	73,257	26,023	3,139

- Training and Augmentation

- 70:30** real-synthetic with pixel-level augmentations, trained for **300 epochs** under identical budgets.

- Evaluation Matrices

- Measured clean-accuracy, robustness to adversarial attacks (FGSM, PGD, BIM, AutoAttack) and measure edge efficiency.

LiteAugNET: Experimental Findings

- We **comprehensively evaluated** LiteAugNET across six complementary research questions:
 - RQ1: Clean accuracy evaluation
 - RQ2: Domain-shift generalization
 - RQ3: Adversarial robustness evaluation
 - RQ4: Semantic alignment diagnostics
 - RQ5: Scalability and efficiency analysis on edge devices
 - RQ6: Ablation studies

LiteAugNET: RQI-Clean Accuracy Evaluation

- Verified that semantic-augmentation improves standard classification accuracy under **identical training budgets**.
- LiteAugNet improves by **+1-2.5%** over standard augmentation across all dataset and backbones.

Backbone	No Aug	Std. Aug	GAN-only	Stable Diff.	LiteAugNET
MobileNetV2	89.4±0.4	92.7±0.3	93.0±0.3	93.4±0.3	94.1±0.3
ResNet18	91.5±0.4	94.4±0.3	94.7±0.3	95.0±0.3	95.8±0.2
ViT-Tiny	88.3±0.5	91.5±0.4	91.9±0.4	92.3±0.3	93.2±0.3
Average	89.9±0.4	92.9±0.3	93.2±0.3	93.6±0.3	94.4±0.3

LiteAugNet: RQ2-Domain Shifts Generalization

- Tested robustness under **common distribution shifts** without **retraining**.
- LiteAugNet improves mean Corruption Accuracy (mCA) by **+4.7%** over standard augmentation and **+4.2%** over GAN-only on CIFAR-10-C dataset.

Method	Noise	Blur	Weather	Digital	mCA
Std. Aug	72.8±0.5	77.5±0.4	81.0±0.4	78.2±0.5	77.4±0.4
GAN-only	73.5±0.5	78.0±0.4	81.5±0.4	78.6±0.5	77.9±0.4
Stable Diff.	74.2±0.5	78.8±0.4	82.2±0.4	79.3±0.4	78.6±0.4
LiteAugNet	77.0±0.4	82.5±0.4	86.0±0.3	82.8±0.4	82.1±0.4

LiteAugNET: RQ3-Adversarial Robustness Evaluation

- Evaluated resistance to **white-box adversarial attacks** without adversarial training.
- LiteAugNet improves robustness by **+5.0-8.2%** across FGSM, PGD, BIM, AutoAttack over Standard Augmentation on CIFAR-10.

Method	FGSM	PGD-20	BIM-10	AutoAttack
Std. Aug	26.0±0.4	7.5±0.4	12.2±0.4	3.5±0.3
GAN-only	26.8±0.4	8.2±0.4	12.9±0.4	3.9±0.3
Stable Diff.	27.5±0.4	9.0±0.4	13.6±0.4	4.4±0.3
LiteAugNet	33.5±0.4	15.7±0.4	19.8±0.3	8.5±0.3

LiteAugNET: RQ4- Semantic Alignment Diagnostics

- Measured whether filtering improves **semantic consistency**, **diversity** and **quality** of synthetic samples.
- The three-stage filtering improves semantic alignment **(+0.12 CLIP)** retaining only high-quality samples.

Dataset	Pre-filter	Post-filter	AR (%)	FID (%)
CIFAR10	0.22 (0.15)	0.34 (0.08)	62	8
Tiny-ImageNet	0.18 (0.18)	0.30 (0.10)	58	6
SVHN	0.25 (0.12)	0.37 (0.07)	64	7

LiteAugNET: RQ5-Scalability and Efficiency Analysis on Edge

- Compared **generation quality vs computational cost** on representative edge devices.
- LiteAugNet achieves real-time augmentation (**≤ 2.5 ms/sample, 38-44 FPS, < 1 GB memory**) reducing energy consumption than GAN-only augmentation.

Edge Device	Method	Gen. Lat (ms)	Throughput	Energy	Memory
Jetson Nano (4GB)	Std. Aug	0.3 ± 0.0	41 ± 1	5.6 ± 0.1	780 ± 8
	GAN (no sem.)	1.8 ± 0.1	34 ± 1	7.9 ± 0.2	810 ± 7
	LiteAugNet	2.1 ± 0.1	38 ± 1	6.5 ± 0.1	830 ± 9
Raspberry Pi 5 (8 GB)	Std. Aug	0.4 ± 0.0	47 ± 1	4.8 ± 0.1	720 ± 7
	GAN (no sem.)	2.2 ± 0.1	41 ± 1	6.5 ± 0.2	770 ± 6
	LiteAugNet	2.5 ± 0.1	44 ± 1	5.5 ± 0.1	780 ± 8

LiteAugNET: RQ6-Ablation Studies

- Identified the contribution of semantic filtering, LoRA adaptation, and adaptive data.
- Removing filtering or LoRA reduces accuracy by +1.9% and robustness by 4.7%.

Variant	Top-1	PGD-20	FID	CLIP	AR (%)
Full LiteAugNet	95.8	15.7	4.2	0.34	62
w/o Filtering	94.3	12.2	4.9	0.26	100
w/o LoRA (linear)	93.9	11.0	4.8	0.29	60
R=0.1 (mix=50%)	95.0	11.5	4.3	0.33	62
R=0.5 (mix=50%)	95.4	14.2	4.7	0.33	62
w/o blending	95.3	14.0	4.2	0.34	62

Key Takeaways

- *LiteAugNet*, the first lightweight semantic-guided augmentation framework, enables real-time edge-level image classification.
- A practical approach to semantic augmentation using semantic-guided synthesis with three-stages on-device filtering pipeline.
- Delivers real-time deployment on edge devices while maintaining accuracy, robustness, and efficiency.

Thank You!

- Webpage: <https://mdshahedrahman.github.io/>

