SEGA-NET: LLM-Guided Semantic-Enhanced GAN Augmentation Network for Low-Resolution Image Classification

Mohammad Shahedur Rahman
UNIVERSITY OF TEXAS ARLINGTON

Mohammad Tahmid Bari
UNIVERSITY OF WATERLOO

Md Delwar Hossain
Angelo State University

Roya Choupani
Angelo State University

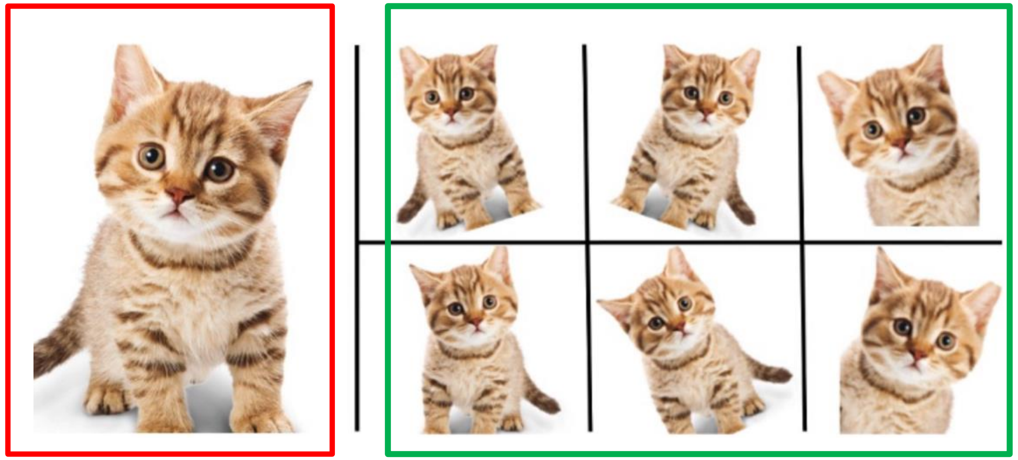
Erdogan Dogdu
Angelo State University

AIxDKE

UNIVERSITY OF TEXAS ARLINGTON

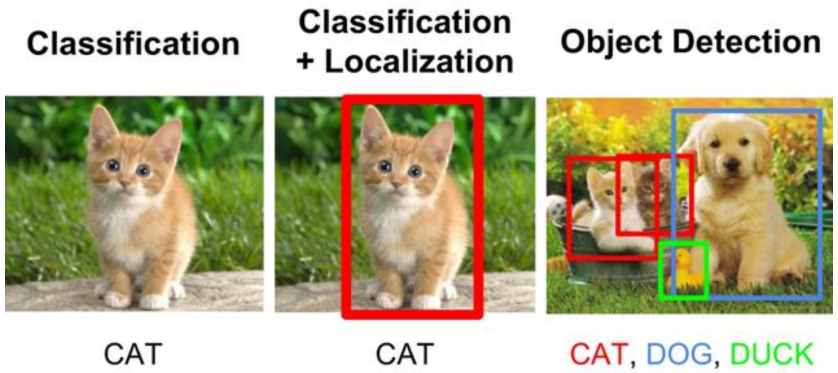
Data Augmentation

- 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 Data Augmentation Matters for Low-Resolution Vision?

- Motivation:
  - Limited semantic capacity and overfitting: it lacks *semantic context*, causing models to overfit to *spurious textures and local patterns.*



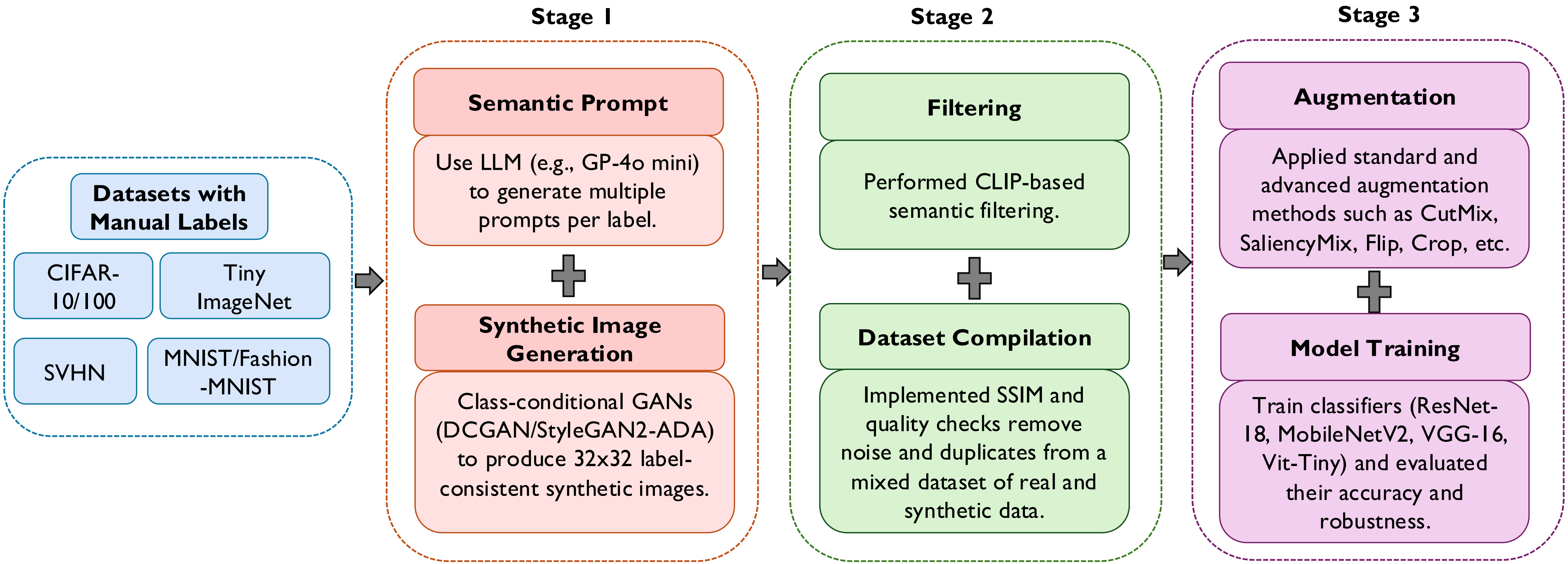
  - Limited effectiveness of standard augmentation: They offer *pixel-level variations only*, failing to introduce *semantic diversity.*
  - Limited robustness and generalization: Models trained on low-resolution data show *degraded robustness* under *distribution shifts and adversarial attacks.*

- *An effective study on data augmentation is required for the low-resolution vision.*

Our Contribution: SEGA-NET

- SEGA-Net: a semantic-aware three-stage augmentation framework for efficient and robust low resolution image classification.
  - Semantic-guided prompt generation
    - *First framework* to *bridge* LLM semantics with class-conditional GANs.
  - Security-aware semantic filtering
    - Apply *CLIP-based alignment, SSIM de-duplication* and *quality-check* to retain label-consistent, high-fidelity synthetic samples.
  - Robust model training
    - Merge filtered synthetic data with real data and apply *pixel-level augmentation*.

Overview of SEGA-NET



SEGA-NET: Semantic-Guided Prompt Generation

- LLM-based expansion
  - Expand each class label into *multiple semantically rich textual prompts* using an LLM (GPT-4o-mini).
- Text-to-latent semantic bridging
  - Map CLIP-embedded prompts into the GAN latent space via a *lightweight text-to-latent projector*.
- Efficient low-resolution synthesis
  - Steer a *frozen class-conditional GAN* to generate *label-consistent 32x32 synthetic images* with low computational cost.

Algorithm 1 (Stage-1: Training the prompt→latent projector P_θ (with G frozen)

Require: Frozen class-conditional GAN G (pretrained per dataset); CLIP encoders (ϕ_v, ϕ_t); class prompts $\{\mathcal{P}_y\}$; clusters $\{C_{y,j}\}$ (from cosine similarity, m=3 fixed); noise scale σ; temperature τ; regularization weight λ; learning rate η; total steps T

Ensure: Projector $P_\theta : \mathbb{R}^{d_{\text{text}}} \to \mathbb{R}^{d_z}$

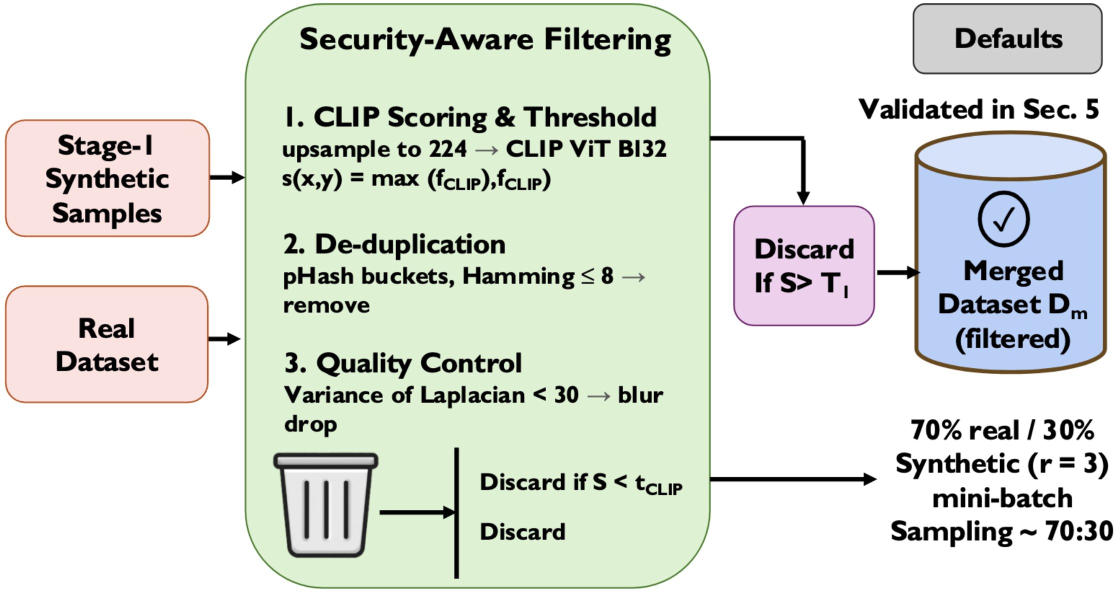
1: Initialize MLP projector $P_\theta : d_{\text{text}} \to 1024 \to d_z$ with GELU and LayerNorm
2: **for** $t = 1$ to T **do**
3: Sample a class y and a cluster $C_{y,j}$; form pair $(\bar{e}_{y,j}, p^+)$ with $p^+ \in C_{y,j}$
4: Compute latent vector: $z \leftarrow P_\theta(\bar{e}_{y,j}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$
5: Generate synthetic image: $x \leftarrow G(z, y)$ ▷ Forward pass only; G frozen
6: Compute CLIP contrastive loss (InfoNCE): $\mathcal{L}_{\text{CLIP}}$
7: Compute regularization term: $\mathcal{L}_{\text{reg}} = \lambda \| P_\theta(\bar{e}_{y,j}) \|_2^2$
8: Update projector parameters:

$$\theta \leftarrow \theta - \eta \nabla_\theta \big(\mathcal{L}_{\text{CLIP}} + \mathcal{L}_{\text{reg}} \big)$$

9: **end for**

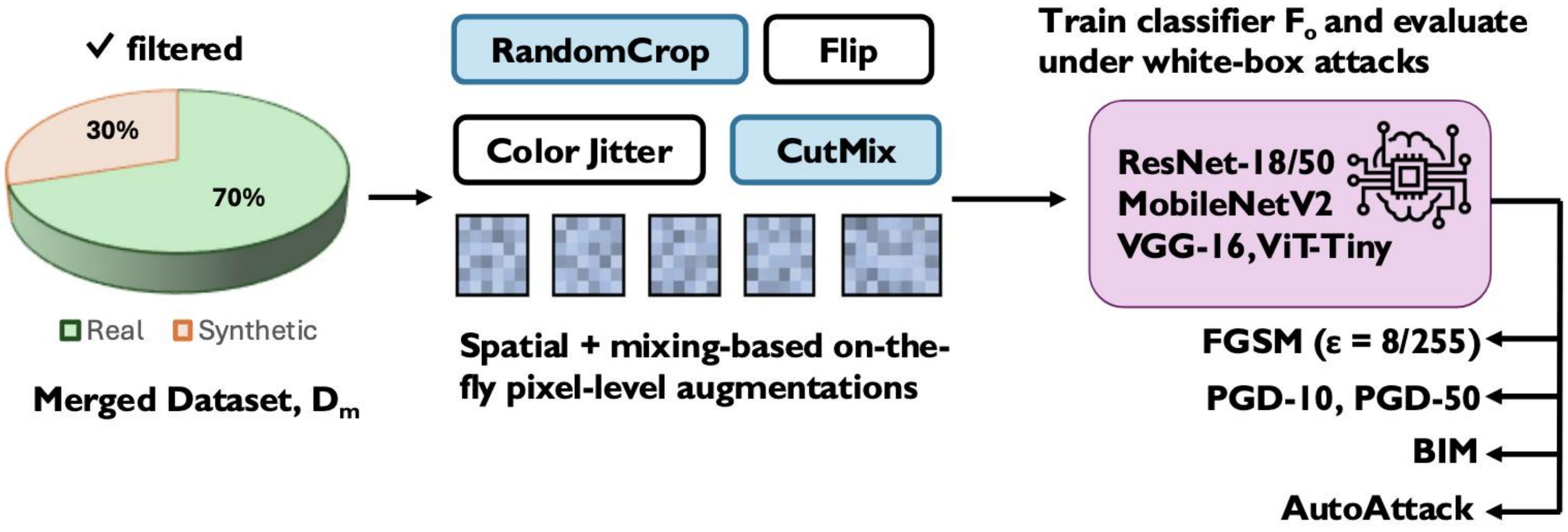
SEGA-NET: Security-Aware Semantic Filtering

- Sematic alignment validation
  - Use *CLIP-based similarity scoring* to ensure generated images align with class semantics.
- Redundancy and quality control
  - Remove near-duplicates using *pHash + SSIM* and discard low-quality samples via *blur detection.*
- Clean synthetic dataset construction
  - Retain only *high-fidelity, label-consistent synthetic samples* for downstream training.



SEGA-NET: Robust Training with Augmentation

- Real-synthetic data integration
  - Merge filtered synthetic samples with real data using a *controlled ratio* to preserve class balance.
- Augmented and robust learning
  - Train classifier with *pixel-based* and *mixing-based* augmentations to improve *accuracy, robustness,* and *generalization.*



SEGA-NET: Experimental Setup

- Dataset and Models
  - CIFAR 10/100, Tiny-Imagenet, Fashion-MNIST (32x32) with ResNet, MobileNetv2, VGG-16 and ViT-Tiny.

Dataset	Cls	Train	Test	N_{syn}
CIFAR-10	10	50,000	10,000	21,428
CIFAR-100	10	50,000	10,000	21,428
Tiny-ImageNet	200	100,000	10k	42,857
SVHN	10	73,257	26,023	31,498
Fashion-MNIST	10	60,000	10,000	25,714

- 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 generalization under distribution shifts.

SEGA-NET: Experimental Findings

- We comprehensively evaluated SEGA-NET across six complementary dimensions:
  - Clean accuracy evaluation
  - Generalization under domain shifts
  - Adversarial robustness evaluation
  - Semantic alignment evaluation
  - Scalability and efficiency analysis
  - Sensitivity and ablation studies

SEGA-NET: Clean Accuracy Evaluation

- Verified that semantic-augmentation improves standard classification accuracy under identical training budgets.
- SEGA-Net improves by +8-10% over standard augmentation across all dataset.

Method	CIFAR10	CIFAR 100	Tiny ImageNet	SVHN	Fashion MNIST
No Aug	68.5±0.6	48.5±0.5	35.2±0.5	83.9±0.4	90.6±0.3
Std. Aug	73.2±0.5	50.4±0.6	36.9±0.5	88.1±0.3	93.2±0.3
GAN-only	74.0±0.4	51.8±0.5	37.2±0.6	89.2±0.4	94.0±0.2
Stable Diff.	74.5±0.5	52.1±0.5	37.8±0.6	90.1±0.3	94.7±0.2
SEGA-NET	**78.3±0.4**	**53.7±0.5**	**39.2±0.6**	**92.0±0.3**	**96.5±0.2**

SEGA-NET: Generalization Under Domain Shifts

- Tested robustness under corruptions and cross-domain-shifts without retraining.

- SEGA-Net improves accuracy by +5.9% over standard augmentation and +10.3% over No augmentation on CIFAR-10-C dataset.

Method	**CIFAR 10-C**	**CIFAR 100-C**	**Tiny IN (C)**	**SVHN MNIST**	**F-MNIST**
No Aug	54.2±0.5	32.1±0.6	21.4±0.4	65.3±0.6	71.2±0.5
Std. Aug	58.7±0.6	35.4±0.5	24.8±0.5	72.0±0.5	76.5±0.4
SEGA-NET	**64.5±0.4**	**39.6±0.5**	**28.2±0.5**	**78.1±0.4**	**82.3±0.3**

SEGA-NET: Adversarial Robustness Evaluation

- Evaluated resistance to white-box adversarial attacks without adversarial training.
- SEGA-Net improves robustness by +7-12% across FGSM, PGD, BIM, AutoAttack.

Method	FGSM	PGD-10	PGD-50	BIM	AutoAttack
No Aug	43.2±0.6	29.4±0.5	25.1±0.5	31.0±0.6	23.5±0.5
Std. Aug	48.7±0.5	33.6±0.6	29.8±0.5	36.2±0.5	27.4±0.4
SEGA-NET	**55.3±0.4**	**41.2±0.5**	**36.5±0.5**	**43.8±0.4**	**34.1±0.4**

SEGA-NET: Semantic Alignment Evaluation

- Measured whether filtering improves semantic consistency of synthetic samples.
- CLIP filtering increases median similarity by +0.10-0.15 across datasets.

Method	CIFAR10	CIFAR 100	Tiny IN	SVHN	F-MNIST
Pre-filter	0.65 / 0.18	0.55 / 0.22	0.50 / 0.25	0.65 / 0.19	0.60 / 0.21
Post-filter	**0.75 / 0.10**	**0.70 / 0.12**	**0.65 / 0.14**	**0.80 / 0.08**	**0.75 / 0.11**

SEGA-NET: Scalability and Efficiency Analysis

- Compared generation quality vs computational cost with diffusion and GAN baselines.
- SEGA-Net is +6.7x more compute efficient than diffusion-based augmentation.

Variant	sec/img	GPU (h)	Mem (GB)	Params (M)	FID	Quality (1-5)
SEGA-NET (full)	0.45	12	16	50.2	12.5	4.3
SEGA-NET (no GAN)	0.42	11.5	15	50.2	14	4
SEGA-NET (no LLM)	0.4	11	15	50.2	13.8	4.1
Stable Diffusion	3.2	80	24	1100	8.2	4.8
Diffusion (no filter)	2.8	75	22	1100	9	4.7

SEGA-NET: Sensitivity and Ablation Studies

- Identified the contribution of LLM prompts, GAN synthesis, and filtering.
- Removing any component reduces accuracy by +3.4% average accuracy.

Method	CIFAR 10	CIFAR 100	Tiny ImageNet	SVHN	Fashion MNIST	Avg.
GAN-only	70.1	45.3	32	89.5	85	64.4
LLM-only	72	46.8	33.5	90.3	86.2	65.8
GAN+LLM	73.5	48.2	35.1	91.2	87.4	67.1
GAN+Filter	71.8	47	34	90	86.8	65.9
LLM+Filter	74.2	49	36.2	92.1	88	67.9
SEGA-NET	**78.3**	**53.7**	**39.2**	**94.5**	**90.6**	**71.3**

Key Takeaways

- *SEGA-Net*, the first semantic-aware GAN augmentation framework integrating LLM-guided prompts for low-resolution image classification.
- A principled approach to semantic augmentation using projector-steered GANs with quality-controlled filtering.
- Deep understanding of how semantic augmentation in low-resolution image classification improves generalization, robustness, and efficiency.

Thank You!

- Graph Lab @ UTA
- Webpage: https://mdshahedrahman.github.io/