[Samsung PRISM] Preliminary Discussion

GEN AI - IMAGE RESIZERS

Team

1. College Professor(s):
   Dr. Ranjeet Kumar - ranjeet.k@vit.ac.in
   Dr. Jayanthi R - jayanthi.r@vit.ac.in
1. Students:
   1. Rasika Phutane - rasika.suresh2022@vitstudent.ac.in
   2. Pojesh Kumar R - pojeshkumar.r2022@vitstudent.ac.in
   3. C R Lavanya - lavanya.2022a@vitstudent.ac.in
2. Department:
   SCOPE & SENSE

Date: 23 Sep 2024

GenAI | Image Resizers

SAMSUNG

Problem Statement

Context

GenAI image generation models are good at generating images in trained resolution (For example 1024*1024). However, this limitation is not desirable in real life situation where we need to have images in different resolutions.
Another aspect is generally images generated are in square format which is not suitable for usage on mobile or laptop having rectangle format predominant.

Statement

Resizers and auto-upscalers (2X, 4X) for the images. Change to landscape and portrait.

Worklet Details

6	4	Gokul Nair +91-9496059290 gokul.gn@samsung.com
Duration (Months)	**Members Count**	Mentors

Pre-Requisite

- https://openmodeldb.info/
- https://paperswithcode.com/task/image-super-resolution

Expectations

Undertaken Tasks

- Conduct Literature survey
- Identify the suitable framework
- Build a framework for image resizing and upscaling

KPI

- Web application with simple UI. ComfyUI is preferred.
- It should seamlessly integrate with the backend GenAI models. SDXL etc.
- Latency should be <10 seconds
- Original image contents should remain constant.
- No visible drop in image quality.

Timeline

Kick Off **< 1st Month >**	**Milestone 1** **< 2nd Month >**	**Milestone 2** **< 6th Month>**
• Problem Briefing • Check Feasibility • Technical survey	• Finalize the model/pipeline • Architecture diagram	• Final product

Complexity

1 2 3 4 5 6 7 8 9 10

Perceptual Losses for Super Resolution

Two approaches were used **feedforward convolutional neural networks** and **perceptual loss functions**.

Feed-forward networks: These networks are trained using a pixel-wise loss (comparing the predicted image with the ground truth on a pixel level).

Challenge faced : **Feed-forward networks** often misses out on high-level features important for image quality.

Perceptual loss functions: Instead of comparing images pixel by pixel, this loss measures the difference in high-level features extracted from a pretrained network (such as a convolutional neural network trained for object recognition). These features capture more semantic information, such as textures, edges, and shapes.

Ground Truth	Bicubic	Ours (ℓ_{pixel})	SRCNN [11]	Ours (ℓ_{feat})
This Image	21.69 / 0.5840	21.66 / 0.5881	22.53 / 0.6524	21.04 / 0.6116
Set14 mean	25.99 / 0.7301	25.75 / 0.6994	27.49 / 0.7503	24.99 / 0.6731
BSD100 mean	25.96 / 0.682	25.91 / 0.6680	26.90 / 0.7101	24.95 / 63.17

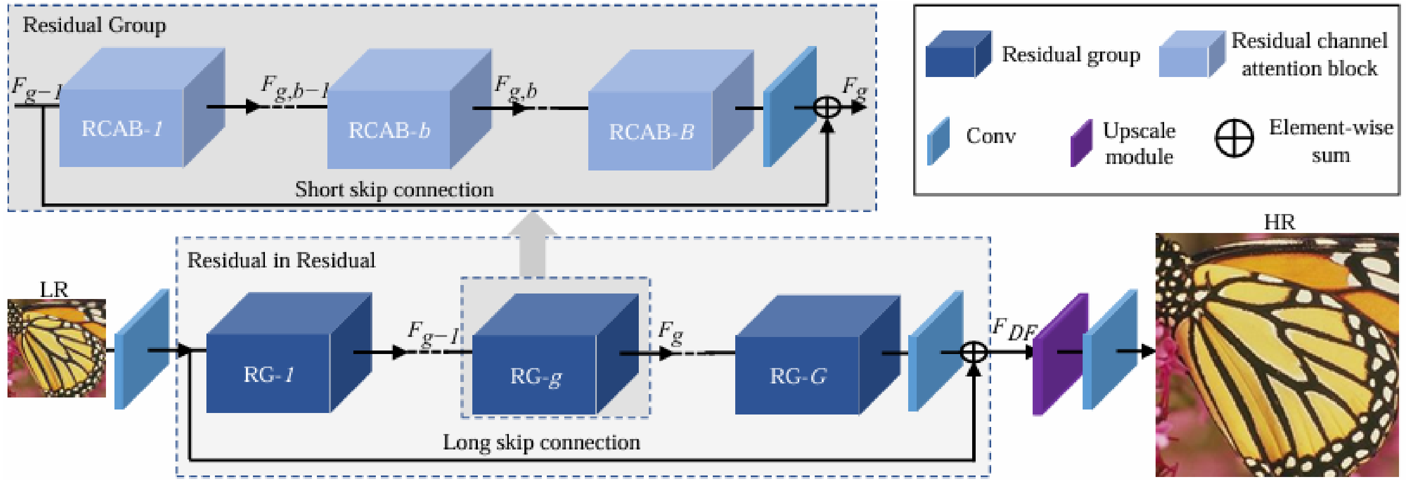
Reference: https://paperswithcode.com/paper/perceptual-losses-for-real-time-style

Residual Channel Attention Network

This model uses a **Residual-In-Residual**(RIR) structure having residual groups with long skip connections and each residual block with short skip connection.

Low frequency information from low-res images bypassed through skip connections while main network handles high frequency information.

Adaptive attention to channel wise features , offer enhanced performance.

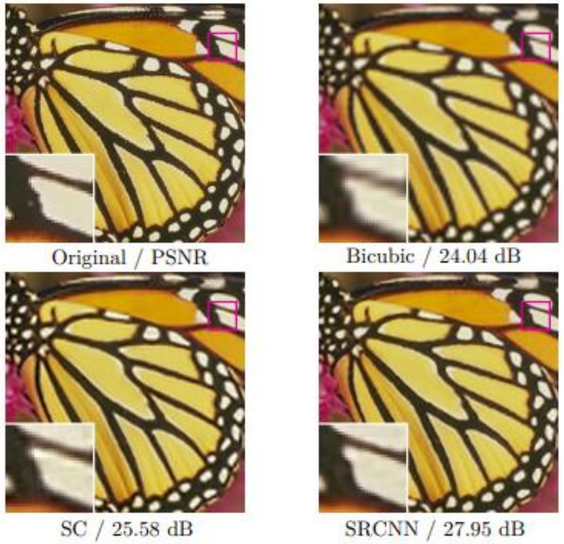


Reference : https://paperswithcode.com/paper/image-super-resolution-using-very-deep

Super-Resolution Using Deep Convolutional Networks

SRCNN takes low-resolution images and directly reconstructs high-resolution versions using a series of convolutional layers, learning to enhance details like edges and textures.

By training on large datasets of LR-HR image pairs, SRCNN learns how to predict missing pixel information, enhancing the details and textures in the output.

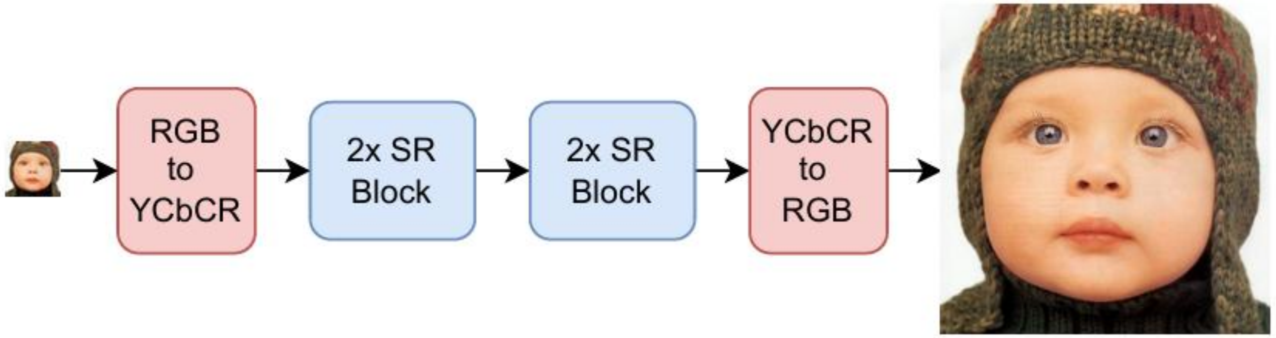


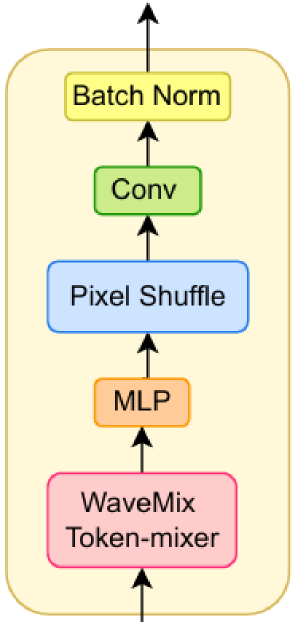
Reference: https://paperswithcode.com/paper/image-super-resolution-using-deep

WaveMixSR - V2

waveMixSR-v2 uses Multilevel wavelet transform layers with pixel shuffling mechanism. This system performs resolution doubling by default and progressive resolution doubling for larger upscaling parameter.

Mode consumes very few resources exhibiting higher parameter efficiency, lower latency and higher throughput.





WaveMixSR-V2

Reference: https://paperswithcode.com/paper/wavemixsr-v2-enhancing-super-resolution-with

Identified Resource Requirements

- Mention any GPU/Device etc requirement here clearly.
- Also mention which of these can be fulfilled at college level (Lab etc).

System Requirements:

Capable GPU for content processing

(can be fulfilled at college level)

Academic Calendar Breaks

- Mention any planned academic activity (Internals/Exams/Holidays/Events/Other Internships etc) in next 6 months which will impact project timeline.
- This helps in setting up the expectation of delivery timeline.

1. 12/10/24 to 20/10/24 - Continuous Assessment 2
2. 18/11/24 to 22/11/24 - Final Assessment Test for LAB courses
3. 25/11/24 to 10/12/24 - Final Assessment Test for Theory courses
4. 22/12/24 to 5/01/25 - Winter Vacation
5. 25/01/25 to 1/02/25 - Continuous Assessment 1
6. 26/02/25 to 01/03/25 - Vibrance (College Event)

Queries

- **Challenges** :
  (Discuss in the form of bullets, what are the next action steps, any road blocks / bottlenecks)

Queries:

Resizer and Upscale is single direction outcomes or bi-directional, Auto Resizing and Up-scaling or User control outcome or Both.

Challenge:

Quality management

Professor Comments

Proposed monthly progress approach

Month 1: Literature Review and Understanding
Month 2: Initial Experimentation
Month 3: Development of Prototyping Tool
Month 4-5: Advanced Techniques Implementation
Month 6: Final Evaluation and Documentation

Latest paper reviewed regarding the problem statement area:

'ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks'

Thank you