Group 7: Class Conditional Image Super-Resolution

Manish Sanjay
190477, 190759 roll no.
{manishy, sanjayso}@iitk.ac.in
Indian Institute of Technology Kanpur (IIT Kanpur)

1 Introduction

Super-resolution has been an attractive research topic for the last 2 decades and has a wide range of applications.

- Surveillance: to detect, identify and perform facial recognition on low-resolution images obtained from security cameras.
- Medical: Capturing high-resolution MRI images can be tricky. Super-resolution can help to generate high-resolution images from LR images.
- Media: Super Resolution can be used to reduce server costs as LR images can be upscaled on the fly.

The super-resolution task is unique in that we often know information about the images that we want to improve. Our rationale is to take advantage of this assumption and devise ways to introduce known information to improve the quality of super-resolution models. Class information in the dataset can be used like wearing glasses, gender info and wearing hat in super-resolution. Our project aims to see how the class conditional information will help in super-resolution or fine-tune some of the missing information during super-resolution. Different sections will cover the following topics:

- Related Work: Literature Review
- Proposed Idea: Our Project Idea
- Methodology: The Project Implementation details
- Results: This will contain both the results and the analysis of our project
- Discussion and Future Work: Future Work to improve results
- Conclusion: Summary
- Individual Contribution: Contribution of the group members

2 Related Work

2.1 SRGAN[1]

Ledig proposed a feed-forward network as the generating function that used a perceptual loss that was the weighted combination of several components. In essence, this paper features a deep residual network with the capability for large upscaling factors ($4\times$) with photo-realistic reconstructions of low-resolution images. SRGAN is generally regarded as the state of the art for super-resolution today.

2.2 SRCGAN[2]

In this model, a classifier is trained on the dataset, based on a set of pre-determined attributes (i.e. gender and eyeglasses) . Next classifier's weights were frozen and imported into the super-resolution GAN model. With this connected classifier, they calculated the class loss and added this to the generator loss term and minimized it with an AdamOptimizer in the same way that a vanilla GAN might operate. They could produce images having marginally better glasses as compared to SRGAN when they used the glasses attribute. For gender-based classifier, they couldn't find significant improvement as compared to SRGAN.

2.3 ESRGAN[3]

In this model they remove all BN layers and replace the original basic block with the proposed Residual-in-Residual Dense Block (RRDB), which combines a multi-level residual network and dense connections. They also introduced a relativistic discriminator, and also developed a more effective perceptual loss by constraining features before activation rather than after activation as practised in SRGAN. They also use network interpolation to remove unpleasant noise in GAN-based methods while maintaining good perceptual quality. They reported that their proposed ESRGAN outperforms previous approaches in both sharpness and details. They also got rid of artefacts and produced natural results.

SRCGAN authors reported that gender is such a broad classification that the model would have trouble learning a specific thing to improve on. Super-resolution models generally look for specific, local information to produce a higher resolution image. Thus we should use specific attributes based classifier. We can also improve our accuracy by improving the architecture of our model

3 Proposed Idea

We propose the following changes to the existing SRCGAN architecture.

- SRCGAN uses Residual block as the basic unit. Instead of Residual block, we are going to use Residual in Residual Dense Block.

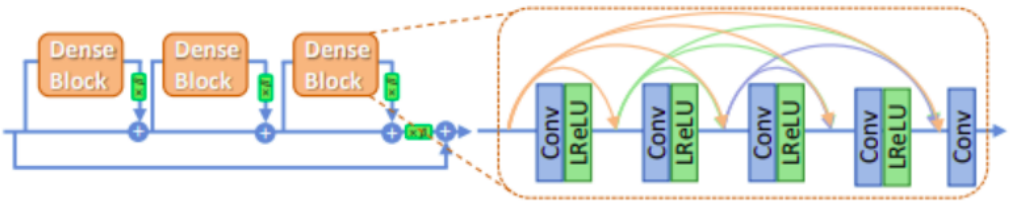


Figure 1: Residual in Residual Dense Block

- We remove the Batch Normalisation Layer instead of this we are going to use residual scaling and smaller initialisation.
- Instead of a simple discriminator, we are going to use a Relativistic discriminator.
- Pre-Activation VGG-Loss: Generally loss is calculated after the activation function but we are going to calculate the loss before activation.
- We also do multi-class Conditional SR with attributes hat and eyeglasses and also on single class conditional SR eyeglasses and gender.

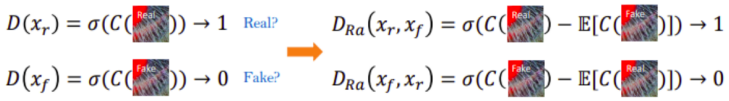


Figure 2: Relativistic Discriminator

We listed the 5 changes above, these are the reason to expect better accuracy by doing these changes:

- Based on the observation that more layers and connections could boost performance, the proposed RRDB employs a deeper and more complex structure than the original residual block in SRCGAN. Specifically, as shown in Fig. 1, the proposed RRDB has a residual-in-residual structure, where residual learning is used at different levels.
- Removing BN layers has proven to increase performance and reduce the computational complexity in Super Resolution and deblurring. BN layers normalize the features using the mean and variance in a batch during training and use the estimated mean and variance of the whole training dataset during testing. When the statistics of training and testing datasets differ a lot, BN layers tend to introduce unpleasant artefacts and limit the generalization ability. It is empirically observed that BN layers are more likely to bring artefacts when the network is deeper and trained under a GAN framework. These artefacts occasionally appear among iterations and different settings, violating the need for stable performance over the training. We, therefore, remove BN layers for stable training and consistent performance. Furthermore, removing BN layers helps to improve generalization ability and to reduce computational complexity and memory usage.
- Use of Relavistic Discrimanator Therefore, our generator benefits from the gradients from both

  Loss of Discriminator in this case

  $$L_D^{Ra} = -\mathbb{E}_{x_r}[\log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(1 - D_{Ra}(x_f, x_r))].$$

  Loss of Generator in this case

  $$L_G^{Ra} = -\mathbb{E}_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(D_{Ra}(x_f, x_r))],$$

  generated data and real data in adversarial training, while in SRCGAN only generated part takes effect.
- Contrary to the convention, we propose to use features to calculate VGG loss before the activation layers, which will overcome two drawbacks of the original design. First, the activated features are very sparse, especially after a very deep network. The sparse activation provides weak supervision and thus leads to inferior performance. Second, using features after activation also causes inconsistent reconstructed brightness compared with the ground-truth image
- As we know that more information we provide to the GAN model will help in super-resolution to fine-tune some of the missing information during super-resolution. Hence with multi-class Conditional SR with attributes hat and eyeglasses, we can expect better results.

These changes are inspired by the papers available on super-resolution, multiple layer residual networks.

4 Methodology

We used celeba dataset for the training all 4 classifiers (Eyeglasses, Gender, Hat, Beard) and all 4 models (ESRGAN, ESRGAN+Gender, ESRGAN+Eyeglasses, ESRGAN+Hat+Eyeglasses

4.1 Data Augmentation

Initial Dimension in Celeba dataset = 178*210
1. First we do centre cropping to 144*144
2. Cropped Image then downsampled to 64*64 which is taken as a high-resolution Image
3. This 64*64 image is further downsampled to 16*16 which is input to the generator.

4.2 Implementation Details

4.2.1 Classifiers Training

For all the classifiers we used Resnet-18 architecture, Adam optimiser, crossentropy as loss function. We also used augmentation like random rotation, changing contrast etc to make classifiers more robust.
1. **Gender Classifier**: As the dataset is gender class balanced. So we use the 200k images as a dataset and then do the train validation test spilt (0.7,0.15,0.15) respectively. We got an accuracy of 94% on the test data.

2. **Eyeglasses Classifier**: For eyeglasses, the dataset is not class balanced. We have only 13k images which have eyeglasses out of 200k. So we take 30k images to balance the dataset (17k without eyeglasses and 13k with eyeglasses) then we do the train validation test split (0.7,0.15,0.15) respectively. We got an accuracy of 93% on the test dataset.

3. **Hat Classifier**: We have 10k images which have hats so we take 25k images (15k without hat and 10k with hat). We do train validation test split as (0.7,0.15,0.15) respectively. we got an accuracy of 93.5% on the test dataset. We used transfer learning for the training of the hat classifier.

4. **Beard Classifier**: We applied the same method for the beard classifier but we got only 75% accuracy. The reason we found out is that in many cases the beard is not in the image due to centre cropping and the blonde beard is also not visible due to downsampling of the images.

4.2.2 GAN Training

1. **ESRGAN**: We take an implementation of ESRGAN and changed the model to fit our input and output dimensions by adding or removing some pooling layers accordingly. Then we trained it on our augmented dataset. As GAN is semi-supervised learning, we don't do a train test split if we want to test the generator ability but if we want to test the discriminator ability then we have to do the split. The paper[2] also didn't split the dataset for GAN training.
Parameters used:
Batch size=16
Epochs=20
Optimizer=Adam
Adam initialised with learning rate=0.002
Input image shape to Generator = 16x16
Output image shape =64x64
2. **ESRGAN+Eyeglasses Classifier**: We integrated the eyeglasses class loss term into the generator loss term by loading the classifier weights and passing the generated image and original High-Resolution image through the classifier then we got the prediction for both the images. Then we calculate the cross-entropy loss for both the predicted values. We treat the original image loss as a label and subtract the loss of generated image. This loss is then backpropagated to the generator to get high-resolution images. The dataset we used is the augmented celeba dataset with 40k images having (13k images with eyeglasses

and 27 k without eyeglasses).
Parameters used:
Batch size=16
Epochs=20
Optimizer=Adam
Adam initialised with learning rate=0.002
Input image shape to Generator = 16x16
Output image shape =64x64
3. **ESRGAN+Gender Classifier**: We integrated the gender class loss term in a similar way as we did in the case of eyeglasses. The dataset is the complete celeba as it is gender class balanced. Parameters used:
Batch size=16
Epochs=20
Optimizer=Adam
Adam initialised with learning rate=0.002
Input image shape to Generator = 16x16
Output image shape =64x64
4. **ESRGAN+Multiple classes Classifiers**: We used 2 attributes Eyeglasses and Hat for multi-class conditional Image Super-Resolution. We loaded both the classifier. The original HR image and Generated image are passed through both the classifier. We calculated the class loss term for each class in a similar way as we did before. Both the class loss term then added and backpropagated to the generator to include the class effect in the generated image. The dataset we used is 40k images (13k having eyeglasses, 10k having a hat, 17k with both) Batch size=16
Epochs=22
Optimizer=Adam
Adam initialised with learning rate=0.002
Input image shape to Generator = 16x16
Output image shape =64x64

5 Results

5.1 Classifier Results

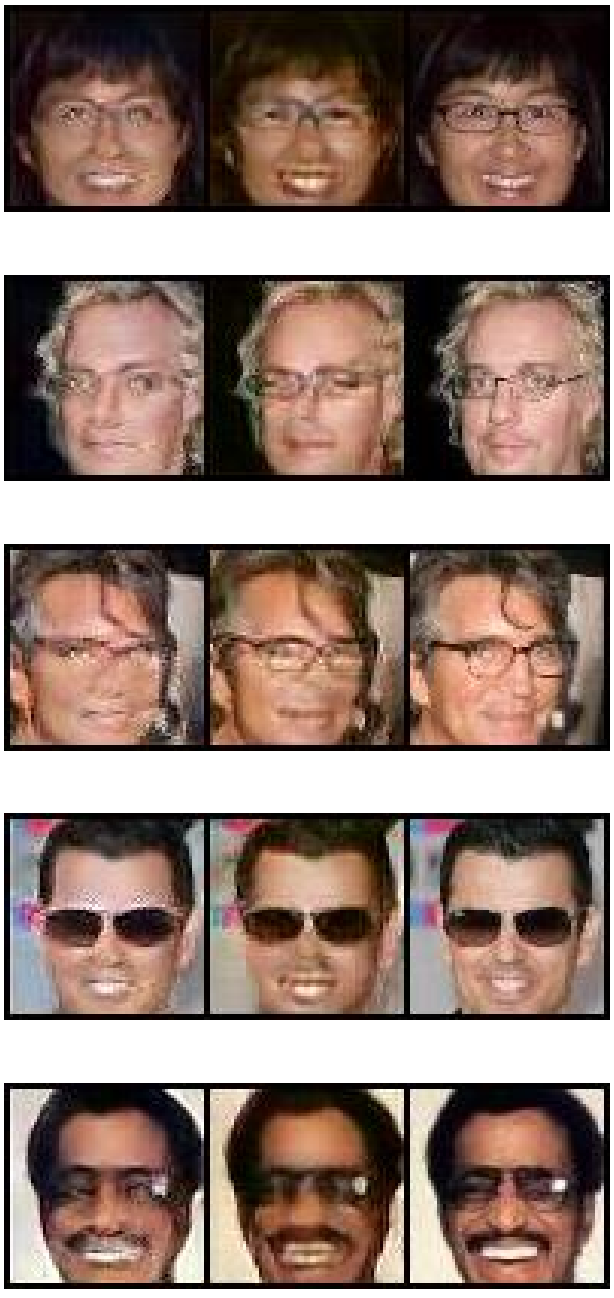
1. **Gender Classifier**: We got an accuracy of 95% on the test dataset with around 94% male class accuracy and 96% female class accuracy.
2. **Eyeglasses Classifier**: We got an accuracy of 94% with 92% eyeglasses class accuracy and 95% without eyeglasses class accuracy.
3. **Hat Classifier**: We got an accuracy of 93.5% with 92% Wearing hat class accuracy and 95% not wearing hat class accuracy.
4. **Beard Classifier**: We got an accuracy of 75% which is very poor and cannot be used with GAN to get better results.

5.2 GAN Results

The paper[2] only shows the images on which improvement was obtained and for some images the predicted class probability but did not analyse the accuracy part of it which we also analyse below and compare the results with ESRGAN to check whether class information was backpropagated or not.
1. **ESRGAN+Eyeglasses Classifier Vs ESRGAN**
Confusion Matrix Comparison:

Result	ESRGAN	ESRGAN+Eyeglasses Classifier
True Positive	7120	10637
True Negative	27982	27563
False Positive	435	854
False Negative	6073	2556
Accuracy	84.35	91.8

Here we can see clearly better theoretical results in the case of the ESRGAN+ Eyeglasses Classifier. Let's see some of the Images on which we get better results.



2. **ESRGAN+ Gender Classifier Vs ESRGAN**
   Confusion Matrix Comparison:

Result	ESRGAN	ESRGAN+Gender Classifier
True Positive	21249	23013
True Negative	16061	15946
False Positive	1766	1881
False Negative	2534	770
Accuracy	89.6	93.62

Here we can also able to see better results than ESRGAN. But in this case, we can only compare the classification accuracy because which images are masculine or feminine changes from person to person but the classifier knows which one is more masculine or feminine and we got better accuracy on ESRGAN + Gender Classifier means Gender information is successfully backpropagated to the generator which now generates better images according to our gender classifier.

3. **ESRGAN+ (Eyeglasses+Hat) Classifier Vs ESRGAN**
   Confusion Matrix Comparison in case of eyeglasses:

Result	ESRGAN	ESRGAN+(Eyeglasses+Hat) Classifier
True Positive	7120	10296
True Negative	27982	27173
False Positive	435	1244
False Negative	6073	2897
Accuracy	84.35	90

Here we can see clear better results in case of accuracy. Let's see some of the images to check whether the images got better or not.

In the above images, we can see better results than ESRGAN. The frame of eyeglasses is crisper in the case of our model.

Figure 3: Leftmost: ESRGAN output, Middle: ESRGAN+ (Eyeglasses +Hat) Classifier, Rightmost: HR Image

The results are consistent on the images having eyeglasses throughout the dataset

Confusion Matrix Comparison in case of Hat:

Result	ESRGAN	ESRGAN+(Eyeglasses+Hat) Classifier
True Positive	9195	9390
True Negative	28994	29340
False Positive	2798	2452
False Negative	623	428
Accuracy	91.77	93

Here We get a little better result than ESRGAN. Let's see some of the images to check whether our model makes hat better or not.

Here also we can clearly see better images in middle than in the left one.

Figure 4: Leftmost: ESRGAN output, Middle: ESRGAN+ (Eyeglasses+Hat) Classifier, Rightmost: HR Image

6 Discussion and Future Work

We divide the analysis of results into 2 sections:

6.1 Single Class Conditional Image Super Resolution

We did 2 experiments on single class. One is ESRGAN+ Eyeglasses Classifier and another one is ESRGAN+ Gender Classifier.

6.1.1 ESRGAN+ Eyeglasses Classifier

In the case of eyeglasses, we got an **7.4%** more accuracy than ESRGAN which is pretty good.
Class wise accuracy:
Wearing Eyeglasses: ESRGAN=54%, Our Model=80.6%
Not Wearing Eyeglasses: ESRGAN=98.4%, Our Model=97%
Here we can clearly see when an image contains eyeglasses our model performed really good in generating eyeglasses than ESRGAN. We got a 26.6% accuracy boost in wearing eyeglasses class.
The class information is getting integrated into the generator by backpropagation. In Training, we can also see the class loss term getting reduced to nearly zero means our model is learning the class aspect of an image and producing a better image by having better eyeglasses. In Figure 3 we can clearly see the better eyeglasses in our model's generated image.

6.1.2 ESRGAN+ Gender Classifier

In the case of Gender, we got an **4%** more accuracy than ESRGAN which is also good.
Class wise accuracy:

Male: ESRGAN=89.3%,Our Model=96.7%
Female: ESRGAN=90%,Our Model=89.4%
Our model performed really good on images labelled as Male. The result we got is vaguer in the case of gender class because after getting generated images from 2 models, it's arguable which one is more masculine or feminine. But the takeaway from it is that classifier plays an important role in integrating class information. What your classifier learns is the same thing the generator will learn and generate better images. So before GAN training just makes sure that the classifier is generalised over different conditions, should be robust and should not be class-biased.

6.2 Multi-Class Conditional Image Super Resolution

For multi-class conditional, we choose eyeglasses and hat as class attributes. We dropped the (beard+mustache) set of attributes as the beard classifier's accuracy is very poor due to centre cropping. Other attributes like chubby, big lips, Narrow eyes, Oval face etc are also not good attributes like gender because with certainty we cannot say which one has better attributes than other image

6.2.1 ESRAGN+ (Eyeglasses+Hat) Classifier

For both classes, we get better results. We got 5.6% better accuracy in the case of eyeglasses and 1.2% in the case of Hat.
Class wise Accuracy:
Wearing Eyeglasses: ESRGAN=54%, Our Model=78%
Not Wearing Eyeglasses: ESRGAN=98.4%, Our Model=95.6%
Here we can see our model performed really good in generating eyeglasses. We got 24% more accuracy for wearing eyeglasses class. We can also see this is the results we got in Figure4. **Class wise Accuracy:**
Wearing Hat:ESRGAN=93.6%, Our Model=95.6%
Not Wearing Hat: ESRGAN=91.1%, Our Model=92.8%
Here we can see a little better result than ESRGAN in both the classes (wearing hat and not wearing hat). In figure 5 also we can see that hat edges and shapes are better in our model than in ESRGAN. Our Model accuracy is upper bounded by hat classifiers accuracy(93.5). If we will increase classifier accuracy means for more images correct class information will be backpropagated. We were not able to increase hat classifier accuracy because in some images centre cropping removed hat.

7 Conclusion

Our project focus on both single class and multi-class conditional image super-resolution. We did 5 changes to the existing class conditional image super-resolution:
1. Uses of Residual in Residual Dense block instead of Residual Block.
2. Use of relativistic discriminator instead of the simple discriminator.
3. Use of Residual Scaling and smaller initialisation instead of Batch Normalisation.
4. Calculation of VGG loss before activation
5. Use of multiple classes for image super-resolution.

We got very good results in the case of eyeglasses. We can also see the class information reflected in the generated images. On gender class, we got good theoretical results but on images, we cannot compare masculinity or femininity. In the case of multiclass conditional, we got good results in the case of eyeglasses class and we can also see this in the generated images. In the case of hat class, we get little better results than ESRGAN which we can also see in the generated images.
Finding and takeaways:
1. Classifier plays an important role in generating better class integrated images. So training of classifiers

should also be given importance. The classifier should be more generalised, and robust and should not be biased toward one class.
2. Choosing attributes also plays a crucial role in how good images you get. One should first run ESRGAN on his/her dataset and then see the generated images. After analysing the missing attributes in the generated images, one should decide the attribute to get better results.
3. Using Multiclass information helped in generating better images having better attributes of both classes. From this we can say after selecting attributes which are missing in generated images, one should train classifiers for that attributes and integrate all the classifiers and train the model. It will help in generating the missing attributes into images.

8 Individual Contributions

Member	Sanjay(190759)	Manish(190477)
Contribution	Project Proposal Project Report Project Presentations Literature Review Gender,Eyeglasses Classifier ESRGAN + Eyeglasses classifier code,training and testing	Project Proposal Project Report Project Presentations Literature Review Gender,Eyeglasses,Hat,Beard Classifier ESRGAN+ Gender Classifier,ESRGAN+ Eyeglasses classifier code,training and testing ESRGAN training and testing ESRGAN+ (Eyeglasses+hat) classifier code,training and testing
% Contribution	40 %	60 %

References

[1] Letdig, "Photo-realistic single image super-resolution using a generative adversarial network.,"

[2] C. W. Vincent Chen, Liezl Puzon, "Class-conditional superresolution with gans,"

[3] S. W. Xintao Wang, Ke Yu, "Esrgan: Enhanced super-resolution generative adversarial networks,"