PCGAN-CHAR: Progressively Trained Classifier Generative Adversarial Networks for Classification of Noisy Handwritten Bangla Characters

Qun Liu, Edward Collier, Supratik Mukhopadhyay

Louisiana State University

Introduction

- There are vast quantities of scanned handwritten text that can be processed and is generally collected as images, which invariably introduces some noise (e.g., damaged documents, noise added due to camera motion, etc.), recognition of handwritten text breaks down with the introduction of noise.
- Most algorithms include denoising step for the images before classification [13], while our approach can directly classify without denoising step due to progressively learn features at increasing resolutions to accurately classify the noisy digits/characters.
- We utilized the progressive technique which is a newly proposed method [14], for training Auxiliary Classifier Generative Adversarial Networks (ACGAN) [20]. It facilitates networks to learn features in a generic to specific manner as the input progresses down the model [23].
- We used the discriminator in the Classifier GAN for our Classification Network for the noisy handwritten characters, and novelly adopted the innovative GAN training technique, Progressive growing [14], to our Classifier GAN.

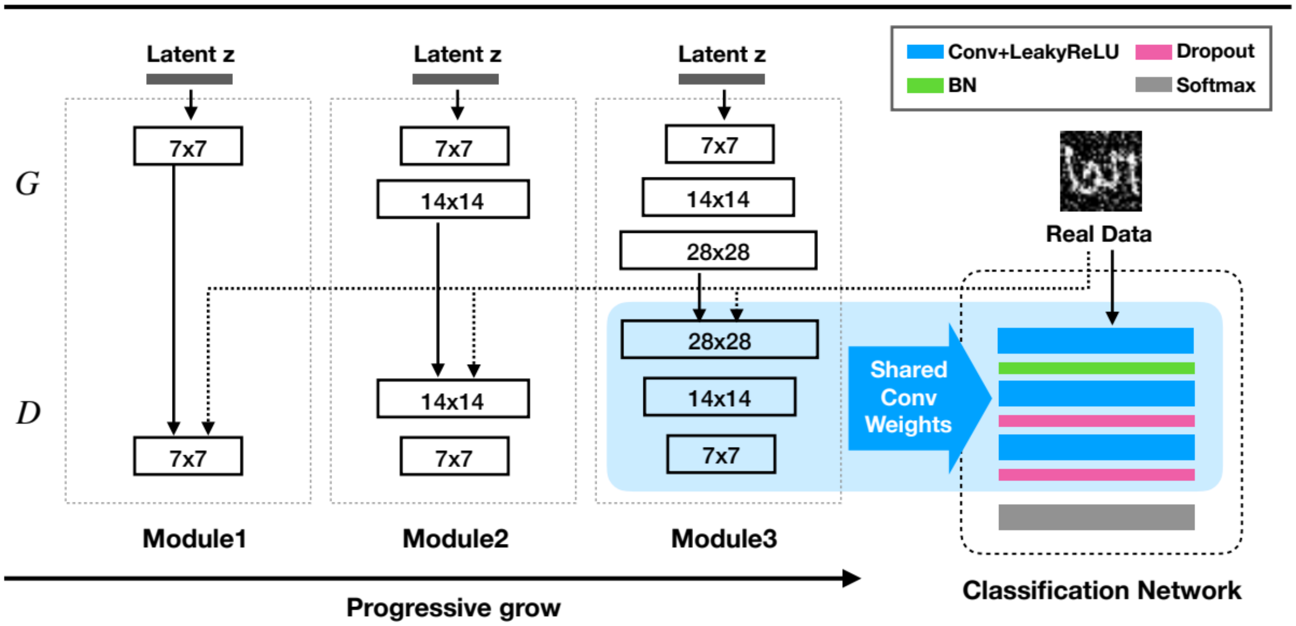
Related Work

- Early work on classification of handwritten characters focused on dimensionality reduction and denoising. This includes the use of quadtrees [1, 3, 18] and intermediate layers of Convolutional Neural Networks for representations and Deep Belief Networks (DBN) for denoising.
- Multi-stage approaches have used chain code histogram features to discriminate classes in [5]. Similar to this work, increasing resolutions are used to assist in classification. Other multistage approaches have used modified quadratic discriminant function (MQDF) and gradients from neural networks to classify characters from many classes [6].
- Auxiliary Classifier GANs (ACGAN) [20] has introduced class labels in GANs by adding an auxiliary classifier which leveraged model in its prior, their research focused on image synthesis tasks and has shown better performance. The focus of our paper is classification of noisy handwritten (Bangla) characters.

Contributions

- It presents a novel robust noise-resilient classification framework using progressively trained classifier general adversarial networks.
- The proposed classification framework can directly classify raw noisy data without any preprocessing steps that include complex techniques such as denoising or reconstruction.
- It experimentally demonstrates the effectiveness of the framework on the Noisy Bangla Numeral, the Noisy Bangla Characters, and the Noisy MNIST benchmark datasets.

Proposed Method



Overview of our framework architecture

Proposed Method

- ***Generative Adversarial Network***: A Generative Adversarial Network (GAN) comprises of two networks: a generator and a discriminator, G plays with D a two-player min/max game, with a loss L(G,D),

$$min_G \ max_D \ \ L(G,D) = \mathbb{E}_y[logD(y)] + \mathbb{E}_{x,z}[log(1 - D(x|G(x|z)))] \tag{1}$$

- ***Classifier GAN***: Comprises of two components each involving two log-likelihood L1 and L2. The L1 involves the conditional probability distribution P(guess|x) where the input x can be a fake image or a real image (a training image from the noisy dataset) and guess takes two values fake or real,

$$L_1 = \mathbb{E}[\log P(real \mid X_r)] + \mathbb{E}[\log P(fake \mid X_f)]. \tag{2}$$

- The log-likelihood L2 involves the conditional probability distribution P(label|x).

$$L_2 = \mathbb{E}[\log P(l \mid X_r)] + \mathbb{E}[\log P(l \mid X_f)]. \tag{3}$$

- The generator G will try to maximize L2 − L1, but the discriminator D will try to maximize L1 + L2.

Proposed Method

Progressively Trained Classifier GAN: Progressive growing is a recent development from [14] that uses transfer learning to improve the quality of the learned models.

- Training is performed individually for the layers of the generator G and the discriminator D.
- New mirroring layers are added to G and D before a new training iteration is run.
- This increases the spatial resolution of the output image for each layer progressively added.
- Layers in G and D become more specialized to spatial resolution, resulting in them learning finer features.

Classification Network: Progressive training results in a stabilized discriminator that has learned the essential features from noisy data at multiple resolutions, resulting in better classification performance. Finally, a softmax layer added as the last layer for classification.

Proposed Method

Algorithm: We focus on classification ability of the discriminator, unlike other variants of GANs focusing on the generator for synthesizing images.

- Initialize the progressive stages
- Initializing a module, begins training the GAN (in normal training procedure).
- An auxiliary classifier added to compute class labels, other than discerning if an image is fake or real. Similar to ACGAN [20].
- After a module is trained, the weights are transferred to the next module.
- Progressive training continues until training for the last module has converged.

Algorithm 1: General Training Algorithm

for *number of modules* **do**
  Initialize(*module*)
  for *epoch=1,2,...,K* **do**
    for *number of batches* **do**
      $B_r \leftarrow$ Sample a batch of n real images with labels
      $B_z \leftarrow$ Sample a batch of n random vectors
      $B_l \leftarrow$ Sample a batch of n random labels
      Update the parameters in the discriminator regarding gradients,

$$\nabla_{\theta_d} \frac{1}{2n} \sum_{r \in B_r, z \in G(B_z, B_l)} \mathcal{L}_{discern}(r, z) + \frac{1}{n} \sum_{r \in B_r} \mathcal{L}_{class}(r)$$

      $B_z \leftarrow$ Sample a batch of $2n$ random vectors
      $B_l \leftarrow$ Sample a batch of $2n$ random labels
      Set discriminator trainable to false, update the parameters in the generator regarding gradients,

$$\nabla_{\theta_g} \frac{1}{2n} \sum_{z, z' \in G(B_z, B_l)} \mathcal{L}_{discern}(z, z') + \frac{1}{2n} \sum_{z, z' \in G(B_z, B_l)} \mathcal{L}_{class}(z, z')$$

    end
  end
  Transfer weights to Next(*module*)
end

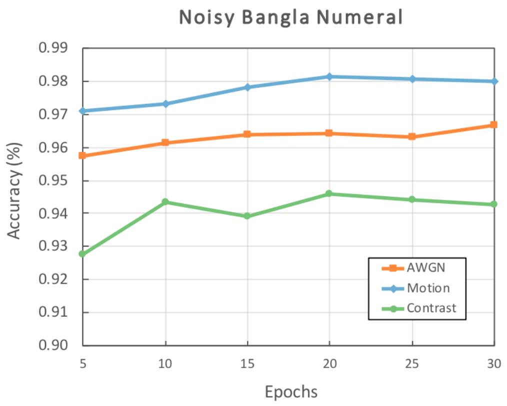
Datasets

- ***Noisy Bangla Numeral***: has three different versions, each with different type of noise added, **AWGN**, **Contrast**, and **Motion**. For each version, there are 10 classes of Bangla Numerals with a total of 23330 black and white images with image size 32 × 32.
- ***Noisy Bangla Characters***: contains 76000 black and white images for 50 classes of Bangla Characters with image size 32 × 32, in each version. There are three different versions one for each type of added noise as stated above.
- ***Noisy MNIST Dataset***: is the same as the original MNIST dataset except for added noise. Again, there are three different versions, one for each of the three types of noise considered. Each version contains 10 classes with a total of 70000 black and white images with image size 28 × 28.

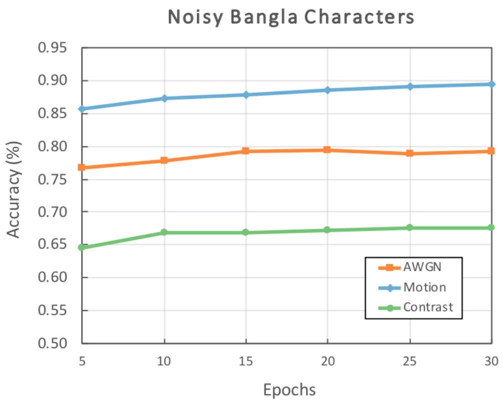
Those data are publicly available datasets could find and download online, it provided a training dataset and a test dataset.

- https://en.wikipedia.org/wiki/List of datasets for machine-learning_research#Handwriting_and_character_recognition

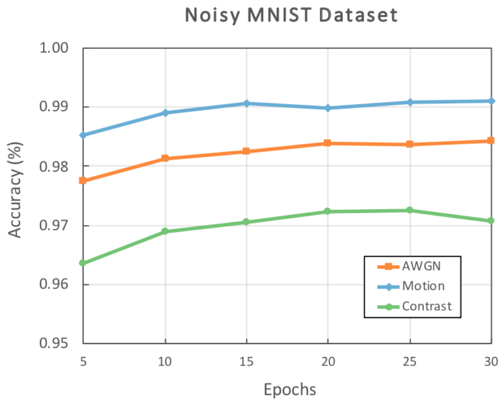
Experiment Evaluation



(a) Classification accuracy on three types of Noisy Bangla Numeral.



(b) Classification accuracy on three types of Noisy Bangla Characters.



(c) Classification accuracy on three types of Noisy MNIST Dataset.

Fig. 3. The classification accuracy of our approach. Our approach has been evaluated on datasets of Noisy Bangla Numeral, Noisy Bangla Characters, and Noisy MNIST with three types of added noise, AWGN, Motion, Contrast.

Experiment Evaluation

Table 1. Comparison of classification accuracy (%) on three types of Noisy Bangla Numeral

Methods	AWGN	Motion	Contrast
Basu et al. [3]	91.34	92.66	87.31
Dropconnect [13]	91.18	97.05	85.79
Karki et al. (w/o Saliency) [13]	95.08	94.88	92.60
Karki et al. (Saliency) [13]	95.46	95.04	92.85
PCGAN-CHAR (Ours)	**96.68**	**98.18**	**94.60**

Table 2. Comparison of classification accuracy (%) on three types of Noisy Bangla characters

Methods	AWGN	Motion	Contrast
Basu et al. [3]	57.31	58.80	46.63
Dropconnect [13]	61.14	83.59	48.07
Karki et al. (w/o Saliency) [13]	70.64	74.36	58.89
Karki et al. (Saliency) [13]	76.74	77.22	**69.66**
PCGAN-CHAR (Ours)	**79.85**	**89.54**	68.41

Table 3. Comparison of classification accuracy (%) on three types of Noisy MNIST dataset

Methods	AWGN	Motion	Contrast
Basu et al. [3]	90.07	97.40	92.16
Dropconnect [13]	96.02	98.58	93.24
Karki et al. (Saliency) [13]	97.62	97.20	95.04
PCGAN-CHAR (Ours)	**98.43**	**99.20**	**97.25**

To understand the statistical significance of the performance improvements obtained by our framework over [13], we used McNemars test (since our framework and [13] had same test datasets). Based on the results of the McNemar tests, the improvements obtained over [13], even in the case of AWGN and contrast variations are statistically significant.

Conclusion

- In this paper, we presented a novel robust noise-resilient classification framework for noisy handwritten (Bangla) characters using progressively trained classification general adversarial networks.
- The proposed classification framework can directly classify raw noisy data without any preprocessing.
- We experimentally demonstrated the effectiveness of the framework on the Noisy Bangla Numeral, the Noisy Bangla Basic Characters, and the Noisy MNIST benchmark datasets.

References

[1] Aref, W.G., Samet, H.: Decomposing a window into maximal quadtree blocks. Acta Informatica 30(5), 425–439 (1993).

[3] Basu, S., Karki, M., Ganguly, S., DiBiano, R., Mukhopadhyay, S., Gayaka, S., Kannan, R., Nemani, R.: Learning sparse feature representations using probabilistic quadtrees and deep belief nets. Neural Processing Letters 45(3), 855–867 (2017).

[5] Bhattacharya, U., Chaudhuri, B.B.: Handwritten numeral databases of indian scripts and multistage recognition of mixed numerals. IEEE transactions on pattern analysis and machine intelligence 31(3), 444–457 (2009).

[6] Bhattacharya, U., Shridhar, M., Parui, S.K., Sen, P., Chaudhuri, B.: Offline recog- nition of handwritten bangla characters: an efficient two-stage approach. Pattern Analysis and Applications 15(4), 445–458 (2012).

References

[13] Karki, M., Liu, Q., DiBiano, R., Basu, S., Mukhopadhyay, S.: Pixel- level reconstruction and classification for noisy handwritten bangla characters. In: 16th International Conference on Frontiers in Handwriting Recognition, ICFHR 2018, Niagara Falls, NY, USA, August 5-8, 2018. pp. 511–516 (2018).

[14] Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. CoRR abs/1710.10196 (2017).

[18] Markas, T., Reif, J.: Quad tree structures for image compression applications. Information Processing & Management 28(6), 707–721 (1992).

[20] Odena, A., Olah, C., Shlens, J.: Conditional image synthesis with auxiliary classifier gans. In: Proceedings of the 34th International Conference on Machine Learning- Volume 70. pp. 2642–2651. JMLR. org (2017).

[23] Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in neural information processing systems. pp. 3320–3328 (2014).

Thank You!