Arabic Handwritten Characters using Deep Transfer Learning

Faculty of Computers and Artificial Intelligence
Helwan University

Author:
Theodoros milad

Abstract

Abstract

Handwritten Arabic Character Recognition (HACR) is a challenging task due to the cursive nature of the script and similarity between characters. This project evaluates four Deep Learning architectures on the **AHCD Dataset**. We compared a custom **VGG-19** trained from scratch against three Transfer Learning models: **ResNet50**, **InceptionV3**, and **MobileNetV2**.

Our results indicate that input resolution plays a critical role. **InceptionV3**, which utilized input upscaling (75×75), achieved the highest accuracy of **80%**. Conversely, models operating on the native 32×32 resolution (ResNet50, MobileNetV2) suffered from information loss due to aggressive spatial pooling, while the deep VGG-19 model failed to converge due to the vanishing gradient problem.

Contents

1 Introduction **3**

2 Methodology **3**

2.1 Data Preprocessing 3
2.2 Model Architectures Evaluated 3

3 Experiments and Results **3**

3.1 Training Dynamics 3
3.2 Quantitative Results 4
3.3 Discussion 4

4 Conclusion **5**

1 Introduction

Arabic is one of the most widely spoken languages, yet Optical Character Recognition (OCR) remains complex. In this project, we utilize the **AHCD Dataset** [1], consisting of 16,800 images of handwritten Arabic letters (28 classes). The goal is to identify the optimal architecture for classifying these low-resolution (32×32) characters.

2 Methodology

2.1 Data Preprocessing

The dataset consists of 32×32 grayscale images. To ensure compatibility with modern architectures, we applied:

- **Normalization:** Scaled pixel values to $[0, 1]$.
- **Grayscale-to-RGB Adapter:** A learnable 3×3 Convolutional layer was used to project the 1-channel input to 3 channels for Transfer Learning models.
- **Resolution Upscaling (Inception Only):** Since InceptionV3 requires a minimum input size of 75×75, we implemented a bilinear resizing layer within the model pipeline. Other models processed the native 32×32 input.

2.2 Model Architectures Evaluated

1. **VGG-19 (Custom):** Trained from scratch [2].
2. **ResNet50:** Transfer Learning with ImageNet weights [3].
3. **MobileNetV2:** Efficient architecture using Depthwise Separable Convolutions [5].
4. **InceptionV3:** Wide architecture using multi-scale processing [4].

3 Experiments and Results

3.1 Training Dynamics

We trained all models for 15-20 epochs. The training curves below (Figure 1) illustrate the validation accuracy over time.

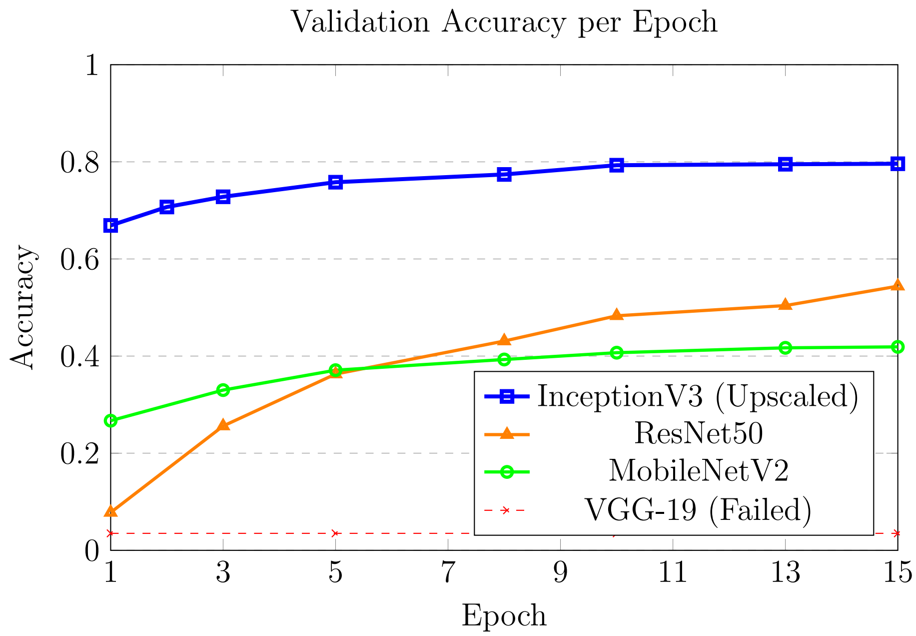

Figure 1: Comparative Training Performance. InceptionV3 demonstrates superior convergence.

3.2 Quantitative Results

The final evaluation metrics on the test set are summarized in Table 1.

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
VGG-19 (Scratch)	3.5%	0.00	0.04	0.00
MobileNetV2	41.9%	0.42	0.42	0.41
ResNet50	54.4%	0.55	0.54	0.54
InceptionV3	**79.6%**	**0.80**	**0.80**	**0.80**

Table 1: Final Test Set Performance Comparison

3.3 Discussion

- **Resolution is Key:** The primary reason for **InceptionV3's** superior performance (80%) is the upscaling of images to 75×75. The cursive nature of Arabic script requires sufficient spatial resolution.

- **Information Loss in Small Inputs:** ResNet50 and MobileNetV2 operated on the native 32×32 images. These architectures use aggressive pooling (downsampling). With such small inputs, the feature maps effectively shrink to 1×1 pixel before classification, causing massive information loss.

- **Vanishing Gradient:** The VGG-19 model (approx 20 layers deep) failed completely without batch normalization or pre-trained weights, proving it is unsuitable for small datasets when trained from scratch.

4 Conclusion

This study concludes that for low-resolution Arabic handwriting, architecture depth is less important than input resolution. **InceptionV3**, combined with image upscaling, is the recommended approach. Future work should investigate upscaling inputs for lightweight models like MobileNetV2 to combine efficiency with accuracy.

References

[1] El-Sawy, A., Loey, M., & EL-Bakry, H. (2017). Arabic Handwritten Characters Recognition using CNN.

[2] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks.

[3] He, K., et al. (2016). Deep Residual Learning for Image Recognition.

[4] Szegedy, C., et al. (2016). Rethinking the Inception Architecture.

[5] Sandler, M., et al. (2018). MobileNetV2: Inverted Residuals.