Fraud Detection

Ideas and Direction

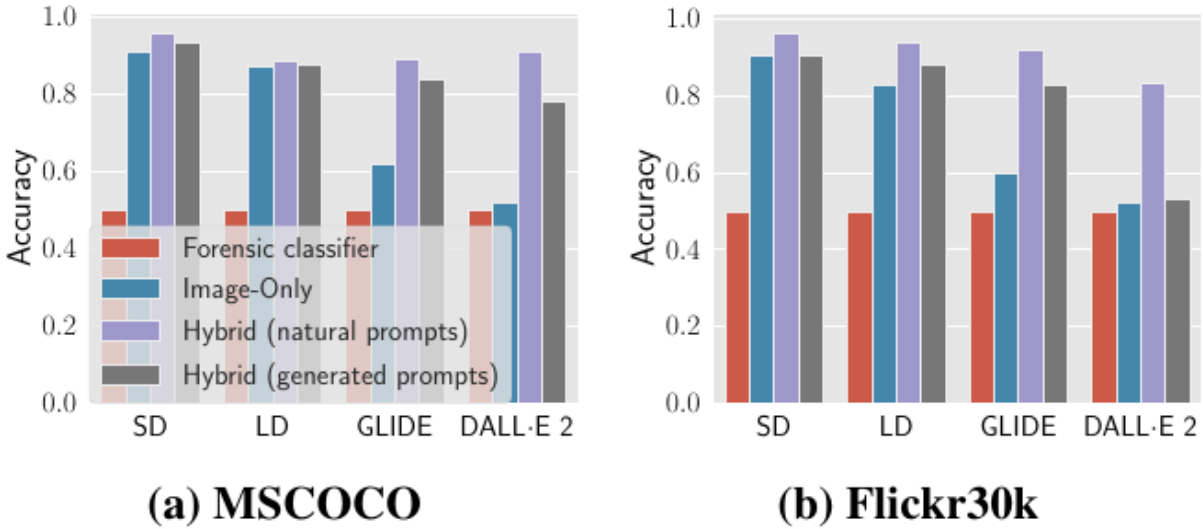
Problem statement

- Classify a given image as real or photoshopped or artificially generated.
- Image Manipulation ways: Photoshopping using Adobe or Adobe Firefly
- Image Generation ways:
  - Text to Image: Diffusion based models for example stable diffusion or DALLE 2.
  - Image to Image: Given an image and a prompt, use diffusion model to manipulate the given image that matches the prompt. For example controlnet based models.

Paper 1: De-Fake by Zeyang Sha and others

Aim: train detection of fake images

- 2 Datasets: MSCOCO and FLICKR. Both has images along with captions.
- Image-only: Run CNN based classifier on images from the dataset (considered real) and the images generated using the captions of the images in dataset, generate fake images using image generation models.
- Hybrid: 2-layer MLP accepting concatenated embedding built from embeddings from Image and its prompt. If prompt not there, BLIP is used.



(a) MSCOCO

(b) Flickr30k

https://arxiv.org/pdf/2210.06998.pdf

Paper 2: Deep Fake Detector

- Used dataset CASIA with 7200 real images and 5331 manipulated images. The manipulation is done using photoshopping or similar tools.
- Input images are analysed using Error Level Analysis that spots compression level differences within an image due to manipulations.
- Using a CNN based classifier on Error Level Analysed images, the authors noted an improvement in the model's performance.

Model architecture	Accuracy on raw	Accuracy on ELA	Precision on ELA	Recall on ELA	F1 Score on ELA
AlexNet	67%	90%	83%	94%	88%
VGG16	68%	88%	82%	89%	86%
Inception	64%	91%	86%	93%	89%
ResNet	75%	94%	92%	93%	93%
DenseNet	67%	89%	94%	95%	89%

Table 1: Comparison of various CNN Models

http://cs230.stanford.edu/projects_winter_2019/reports/15812450.pdf

Research Questions

1. How does CNN based on ResNet architecture trained on personal dataset analysed on Error Level perform compared to the model in paper 2?
2. How does the 2-layered MLP model receiving the ELA processed image and BLIP created textual prompt perform as compared to the model proposed in paper 1.
3. If available the pretrained models from paper 1 and paper 2, compare the performance of detection of fake images by these models on our data against the judgement of the craftmen at Propertyexpert.