**LITERATURE SURVEY**

**TITLE:** Joint face detection and alignment using multitask cascaded convolutional networks

**ABSTRACT:** Face detection and alignment in unconstrained environment are challenging due to various poses, illuminations, and occlusions. Recent studies show that deep learning approaches can achieve impressive performance on these two tasks. In this letter, we propose a deep cascaded multitask framework that exploits the inherent correlation between detection and alignment to boost up their performance. In particular, our framework leverages a cascaded architecture with three stages of carefully designed deep convolutional networks to predict face and landmark location in a coarse-to-fine manner. In addition, we propose a new online hard sample mining strategy that further improves the performance in practice. Our method achieves superior accuracy over the state-of-the-art techniques on the challenging face detection dataset and benchmark and WIDER FACE benchmarks for face detection, and annotated facial landmarks in the wild benchmark for face alignment, while keeps real-time performance.

**TITLE:** "Inceptionism: Going deeper into neural networks."

**ABSTRACT:** [Artificial Neural Networks](http://en.wikipedia.org/wiki/Artificial_neural_network) have spurred remarkable recent progress in [image classification](http://googleresearch.blogspot.com/2014/09/building-deeper-understanding-of-images.html#uds-search-results) and [speech recognition](https://www.youtube.com/watch?v=yxxRAHVtafI). But even though these are very useful tools based on well-known mathematical methods, we actually understand surprisingly little of why certain models work and others don’t. So let’s take a look at some simple techniques for peeking inside these networks.  
  
We train an artificial neural network by showing it millions of training examples and [gradually adjusting the network parameters](https://en.wikipedia.org/?title=Backpropagation) until it gives the classifications we want. The network typically consists of 10-30 stacked layers of artificial neurons. Each image is fed into the input layer, which then talks to the next layer, until eventually the “output” layer is reached. The network’s “answer” comes from this final output layer.

**TITLE:** "Deep fake detection using neural networks."

**ABSTRACT:** Deepfake is a technique for human image synthesis based on artificial intelligence. Deepfake is used to merge and superimpose existing images and videos onto source images or videos using machine learning techniques. They are realistic looking fake videos that cannot be distinguished by naked eyes. They can be used to spread hate speeches, create political distress, blackmail someone, etc. Currently, Cryptographic signing of videos from its source is done to check the authenticity of videos. Hashing of a video file into fingerprints (small string of text) is done and reconfirmed with the sample video and thus verified whether the video is the one originally recorded or not. However, the problem with this technique is that the fingerprints and hashing algorithms are not available with common people. In this paper the proposed system follows a detection approach of Deepfake videos using Neural Networks. Binary classification of deepfakes was done using combination of Dense and Convolutional neural network layers. It was observed that 91% accuracy was obtained in Adam and 88% was obtained in sgd(stochastic gradient descent) for categorial cross entropy. In binary cross entropy, 90% accuracy was seen in Adam and 86% accuracy was noticed in sgd whereas, 86% accuracy in Adam and 80% accuracy in sgd was obtained in mean square.

**TITLE:** "An image is worth 16x16 words: Transformers for image recognition at scale."

**ABSTRACT:** While the Transformer architecture has become the de-facto standard for natural language processing tasks, its applications to computer vision remain limited. In vision, attention is either applied in conjunction with convolutional networks, or used to replace certain components of convolutional networks while keeping their overall structure in place. We show that this reliance on CNNs is not necessary and a pure transformer applied directly to sequences of image patches can perform very well on image classification tasks. When pre-trained on large amounts of data and transferred to multiple mid-sized or small image recognition benchmarks (ImageNet, CIFAR-100, VTAB, etc.), Vision Transformer (ViT) attains excellent results compared to state-of-the-art convolutional networks while requiring substantially fewer computational resources to train.

**TITLE:** "A deep learning approach to universal image manipulation detection using a new convolutional layer."

**ABSTRACT:** When creating a forgery, a forger can modify an image using many different image editing operations. Since a forensic examiner must test for each of these, significant interest has arisen in the development of universal forensic algorithms capable of detecting many different image editing operations and manipulations. In this paper, we propose a universal forensic approach to performing manipulation detection using deep learning. Specifically, we propose a new convolutional network architecture capable of automatically learning manipulation detection features directly from training data. In their current form, convolutional neural networks will learn features that capture an image’s content as opposed to manipulation detection features. To overcome this issue, we develop a new form of convolutional layer that is specifically designed to suppress an image’s content and adaptively learn manipulation detection features. Through a series of experiments, we demonstrate that our proposed approach can automatically learn how to detect multiple image manipulations without relying on pre-selected features or any preprocessing. The results of these experiments show that our proposed approach can automatically detect several different manipulations with an average accuracy of 99.10%.