**Deepfake detection**

Deepfakes make minimal changes to the structure of the faces making them hard to detect. So trying to train models to detect minute changes often leads to overfitting in almost every case.

**Deepfake techniques:**

1. **Attribute Manipulation:** this manipulation, also known as face editing or face retouching, consists of modifying some attributes of the face such as the color of the hair or the skin, the gender, the age, adding glasses, etc
2. **Face2face/identity swap:** Can swap faces of people making them change appearance.
3. **Deepkfakes:** Manipulates facial features according to a given reference.Can be used to orchestrate fake videos of politicians, etc.
4. [**Synthesizing Obama**](https://grail.cs.washington.edu/projects/AudioToObama/) **:** A more advanced version of deepfake.Can manipulate facial expressions according the given input audio sequence
5. **Entire Face Synthesis:** this manipulation creates entire non-existent face images, usually through powerful GAN

**Solutions:**

1. **XceptionNet**

Has shown reliable results in identifying forged videos.Instead of using a single CNN, it makes use of inception model that decouples different tasks that can operate in non overlapping features and concatenate the results in the end.

1. **Face augmentation using face landmark information.**

Video preprocessing method, showed significant gain in performance for deepfake detectors,focuses on extracting face features in order to train model and avoiding overfitting by using face-cutout(Pixel erasing technique that repaints a group of pixels in different shapes using face landmark information)

1. **EfficientNet (EffNet):**

Focuses on efficiently scaling convolution networks to improve performance and speed while keeping the size minimum. EfficientNet-B7 performed 6x times faster while being 8.4x times smaller than a traditional ConvNet

1. **MesoNet:**

There are two architectures in this: Meso-4 and MesoInception4. They are based on well-performing networks for image classification that alternate layers of convolutions and pooling for feature extraction and a dense network for classification.

Meso-4: This network begins with a sequence of four layers of successive convolutions and pooling, and is followed by a dense network with one hidden layer. To improve generalization, the convolutional layers use ReLU activation function that introduce non-linearities and Batch Normalisation to regularize their output and prevent the vanishing gradient effect, and the fully-connected layers use Dropout to regularize and improve their robustness. There are 27,977 trainable parameters for this network.

MesoInception-4: This has been created by replacing the first two convolutional layers of Meso4 by a variant of the inception module. The idea of the module is to stack the output of several convolutional layers with different kernel shapes and thus increase the function space in which the model is optimized. This network has 28,615 trainable parameters overall.

**Feature recognition techniques:**

| **Methods** | **Classifiers/Techniques** | **Key Features** | **Dealing with** |
| --- | --- | --- | --- |
| Eye blinking | LRCN | Use LRCN to learn the temporal patterns of eye blinking.  Based on the observation that the blinking frequency of deepfakes is much smaller than normal. | Videos |
| Intra-frame and temporal inconsistencies | CNN and LSTM | CNN is employed to extract frame-level features, which are distributed to LSTM to construct sequence descriptor useful for classification. | Videos |
| Spatio-temporal features with LSTM | Convolutional bidirectional recurrent LSTM network | An XceptionNet CNN is used for facial feature extraction while audio embeddings are obtained by stacking multiple convolution modules.  Two loss functions, i.e. cross-entropy and KullbackLeibler divergence, are used. | Videos |
| Defenses against adversarial perturbations in deep fakes | VGG and ResNet | Introduce adversarial perturbations to enhance deep fakes and fool deep fake detectors.  Improve accuracy of deepfake detectors using Lipschitz regularization and deep image prior techniques. | Images |
| Phoneme-viseme mismatches | CNN | Exploit the mismatches between the dynamics of the mouth shape, i.e. visemes, with a spoken phoneme.  Focus on sounds associated with the M, B and P phonemes as they require complete mouth closure while deep fakes often incorrectly synthesize it. | Videos |

**References :**

* Improving deepfake detection using dynamic face augmentation: <https://arxiv.org/pdf/2102.09603v1.pdf>
* XceptionNet:<https://ieeexplore.ieee.org/document/8099678>
* DeepFakes and Beyond - A Survey of Face Manipulation and Fake Detection:<https://arxiv.org/pdf/2001.00179v3.pdf>
* FaceForensics++ - Learning to Detect Manipulated Facial Images: <https://arxiv.org/pdf/1901.08971v3.pdf>
* EfficentNet - Rethinking Model Scaling for Convolution Neural Networks:<https://arxiv.org/pdf/1905.11946.pdf>
* MesoNet - a Compact Facial Video Forgery Detection Network: <https://arxiv.org/pdf/1809.00888v1.pdf>