Data Preprocessing

In [1]:
```
import cv2
import os
import pandas as pd
import numpy as np
import shutil
```

In [2]:
```
# shutil.rmtree(os.path.join(os.getcwd(),"2015","preprocessed_data"))
```

In [3]:
```
os.makedirs(os.path.join(os.getcwd(),"2015","preprocessed_data","Training"))
os.makedirs(os.path.join(os.getcwd(),"2015","preprocessed_data","Testing"))
```

In [4]:
```
def create_label(fold):
    label = 0
    if fold == 'Fake':
        label = 0
    else:
        label = 1
    return label
```

In [5]:
```
def folder_struct(input_dir,output_dir):
    imgs, labels=[],[]
    df = pd.DataFrame()
    # Loop through the live and fake subdirectories
    for subfolder in ['Live', 'Fake']:
        # Create the output subdirectory
        os.makedirs(os.path.join(output_dir, subfolder), exist_ok=True)

        # Get a list of image filenames
        image_files = os.listdir(os.path.join(input_dir, subfolder))

        # Loop through the image files
        for image_file in image_files:
            # Read the input image
            image_path = os.path.join(input_dir, subfolder, image_file)
            image = cv2.imread(image_path)

            if image is not None:
                # Normalize the pixel values to the range [0, 1]
                image = 255.0 - image
                image = cv2.resize(image,(300,300))
                # Save the preprocessed image to the output directory
                output_path = os.path.join(output_dir, subfolder, image_file)
                imgs.append(output_path)
                labels.append(create_label(subfolder))
                cv2.imwrite(output_path, image)
            else:
                print(f"Could not read image {image_file}")
    df['Image'] = np.array(imgs)
    df ['Label'] = np.array(labels)
    return df
```

In [6]:
```
input_dir = os.path.join(os.getcwd(),"2015","Training","Digital_Persona")
output_dir = os.path.join(os.getcwd() ,"2015","preprocessed_data","Training")
train_df = folder_struct(input_dir,output_dir)
train_df
```

Out[6]:

	Image	Label
0	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
1	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
2	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
3	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
4	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
...	...	...
13745	/blue/eel6825/ravipatim/2015/preprocessed_data...	0
13746	/blue/eel6825/ravipatim/2015/preprocessed_data...	0
13747	/blue/eel6825/ravipatim/2015/preprocessed_data...	0
13748	/blue/eel6825/ravipatim/2015/preprocessed_data...	0
13749	/blue/eel6825/ravipatim/2015/preprocessed_data...	0

13750 rows × 2 columns

In [7]:
```
train_df.to_csv('Train.csv',index=False)
```

In [8]:
```
input_dir = os.path.join(os.getcwd(),"2015","Testing","Digital_Persona")
output_dir = os.path.join(os.getcwd() ,"2015","preprocessed_data","Testing")
val_df = folder_struct(input_dir,output_dir)
val_df
```

Out[8]:

	Image	Label
0	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
1	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
2	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
3	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
4	/blue/eel6825/ravipatim/2015/preprocessed_data...	1
...	...	...
1245	/blue/eel6825/ravipatim/2015/preprocessed_data...	0
1246	/blue/eel6825/ravipatim/2015/preprocessed_data...	0
1247	/blue/eel6825/ravipatim/2015/preprocessed_data...	0
1248	/blue/eel6825/ravipatim/2015/preprocessed_data...	0
1249	/blue/eel6825/ravipatim/2015/preprocessed_data...	0

1250 rows × 2 columns

In [9]:
```
val_df.to_csv('Val.csv',index=False)
```