# Using Convolutional Neural Networks to Identify Skins in Video Games

Maximilian Phillips – Bangor University School of Computer Science and Electronic Engineering mxp18rrz@bangor.ac.uk

Project Advisor - Prof Ludmila Kuncheva Bangor University School of Computer Science and Electronic Engineering

## Introduction:

There are masses of video data being created every second from video games by the public on streaming services like Twitch. However, there is very little that can be garnered from it in its raw state. The overall aim of this work is to utilise convolution neural networks (CNN) to explore a way in which this data can be turned into something quantifiable.

**Objectives of this project:**

- Explore the probability of classifying a skin with the use of a CNN.
- Experiment to see whether the accuracy can be improved with parameter adjustments or image pre-processing.
- Explore how possible it would be to detect instances of a skin in a video clip.

## Methods:

All experiments use a CNN at the core. So far the dataset has been split 3 ways:
- Resubstituion – all of the data is used for both the training and testing of the network.
- Leave-one-out – each video is in turn left out of the training data to be then used as the testing data.
- Hold-out method – images are split evenly and randomly between testing and training data.

In an attempt to help mitigate the problem of the small dataset a Gaussian filter is utilised. This small amount of pre-processing helps create a more generalised pattern for each skin by blurring it slightly.

It is important to specify the difference between classification, localisation, and object recognition:
- Classification is the ability to say that an image contains a certain object.
- Localisation then adds a bounding box to signify where on this image that object is located.
- Object recognition is being able to detect multiple instances of one or more objects within an image or video.

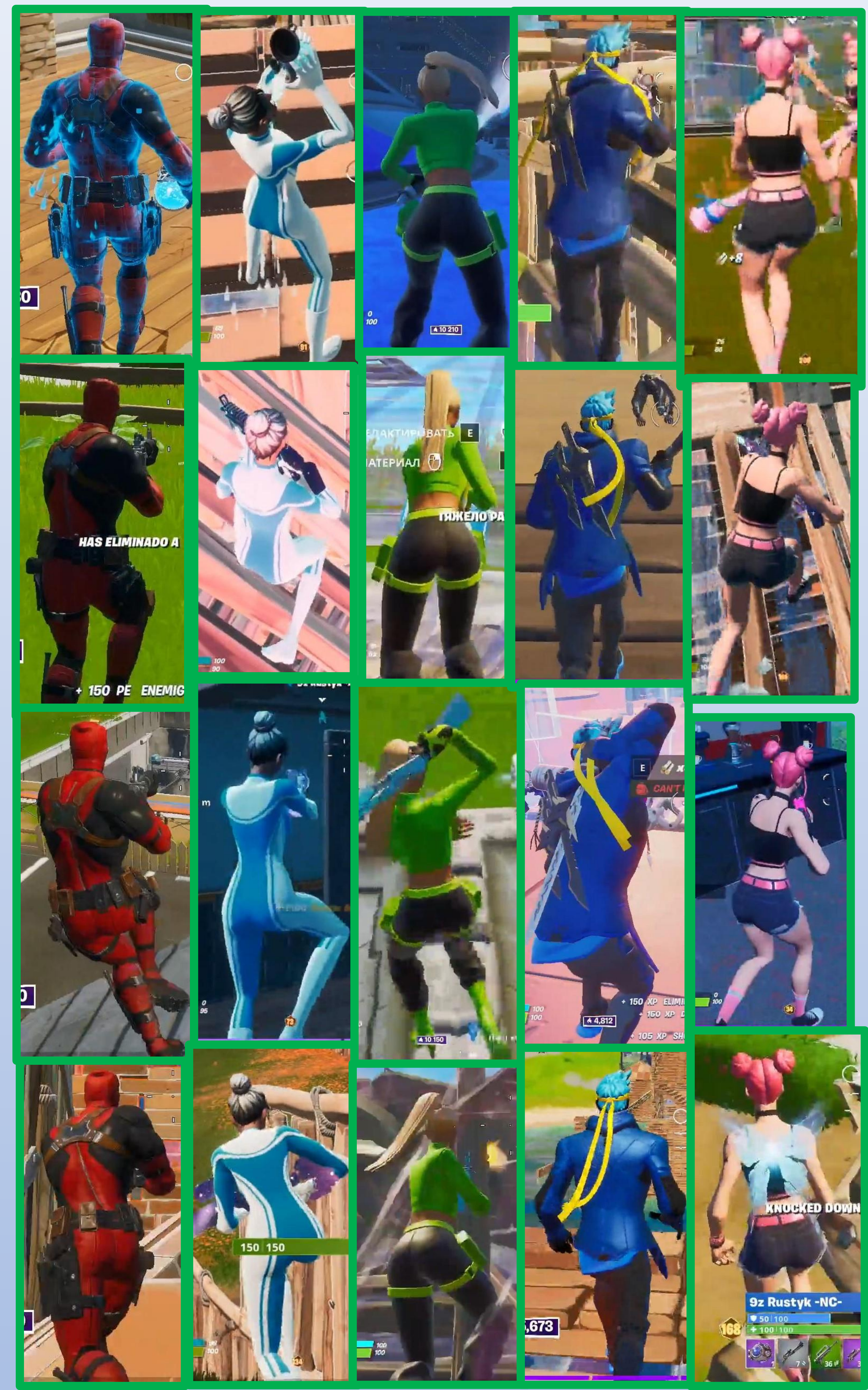


**Fig 1 – A single crop of each video in the dataset**

## Conclusion and Future Work:

Currently, object classification is resulting in good outcomes. However, localisation of an object in an image is computationally expensive. This is due to having to cycle through every potential location of the image for the object. The next step would be to implement the use of the 'Faster R-CNN' method for object recognition in videos..

## Data:

This data set was created from grabbing clips from Twitch. Four videos for five different skins were selected. Each frame of each video was then split into separate images (see fig 1). Crops of the skin in each image were created while saving the coordinates for the bounding box.

To prove that whether skin or no skin could be detected, I first needed gather crops of non-skins to compare the first crops. A short piece of code ran through each frame from every video and cropped an image with randomly generated x and y coordinates. The size of crops was determined by the average size of all of the skin crops.

Even if some of these crops happen to be of the character skin it will be okay as the vast majority won't be. A new CNN was then trained the labels 'skin' and 'notSkin'.

## Experiments:

**Skin type classification:**

Resubstitution test– biased due to the same data appearing on both the training and testing data.
Accuracy – 99.8%

Leave-one-out:
Accuracy – 74.13%

Leave-one-out test – epochs reduced by 50% :
Accuracy – 73.08%

**Skin vs not skin classification:**

Later work will be for exploring potential object recognition and this is the early work for that.

Hold-out test on skin vs not skin data – biased due to crops from the same video appearing in both training and testing data:
Accuracy – 98.29%