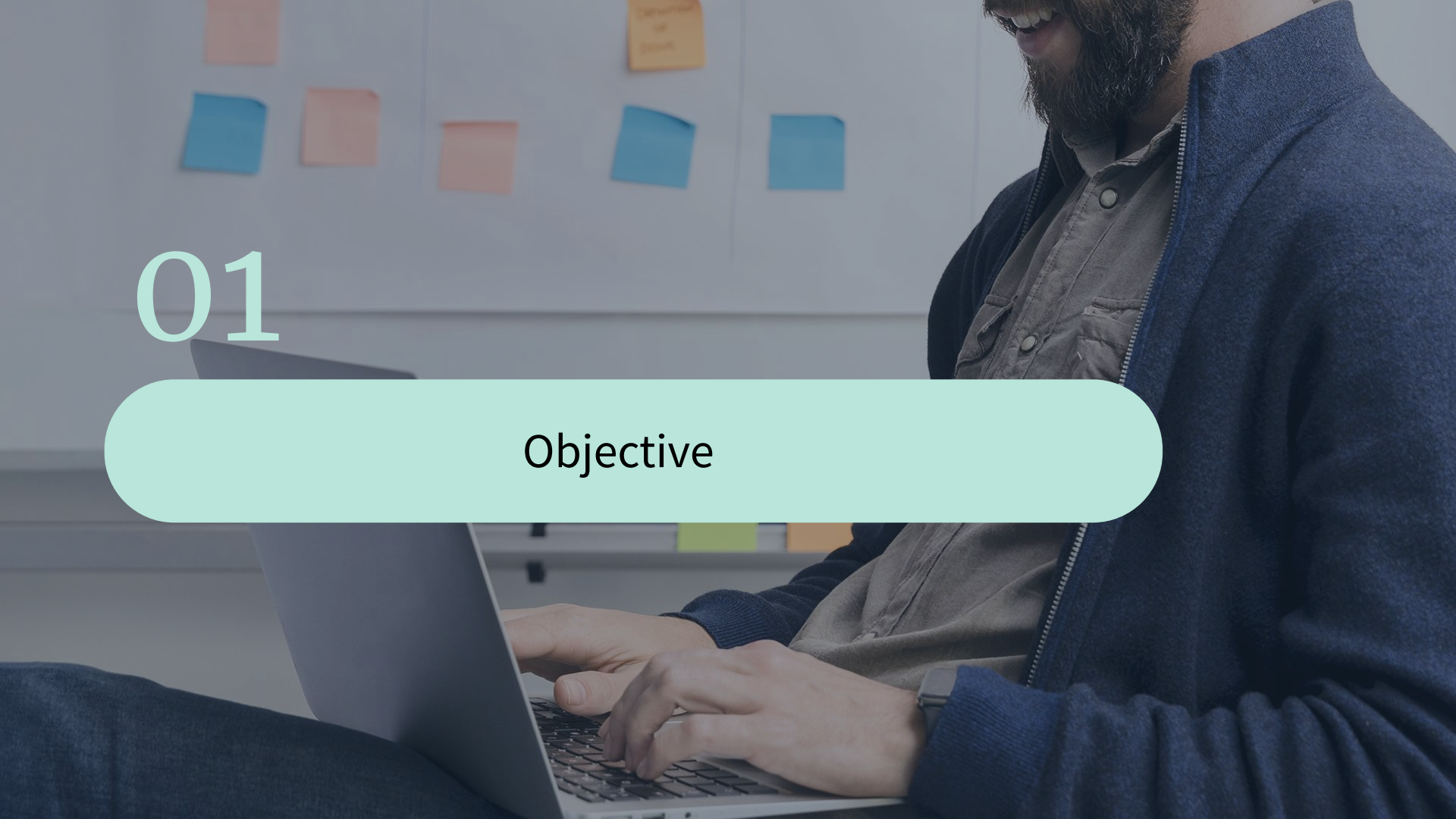
Rich Feature Hierarchies

Object Detection, and Semantic Segmentation

Team reVision | CSE 478 Computer Vision Course Project Spring 2021

Team Members

- Dolton Fernandes
- George Tom
- Naren Akash R J
- Shodasakshari Vidya

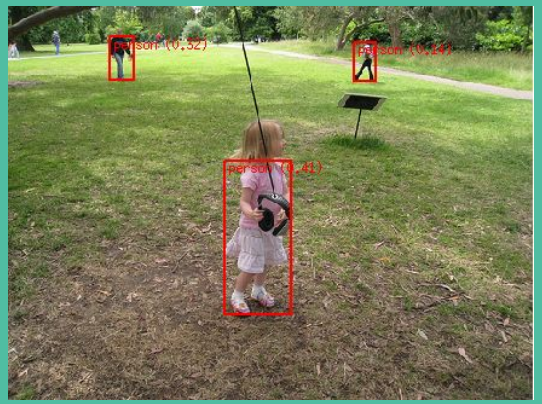


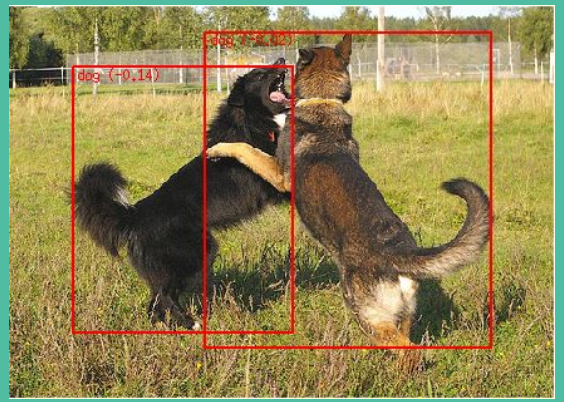
01

Objective

Objective

The objective of the project is to do object detection on images by combining region proposal with CNN.
Given an image, we should be able to identify image regions and the corresponding objects correctly.





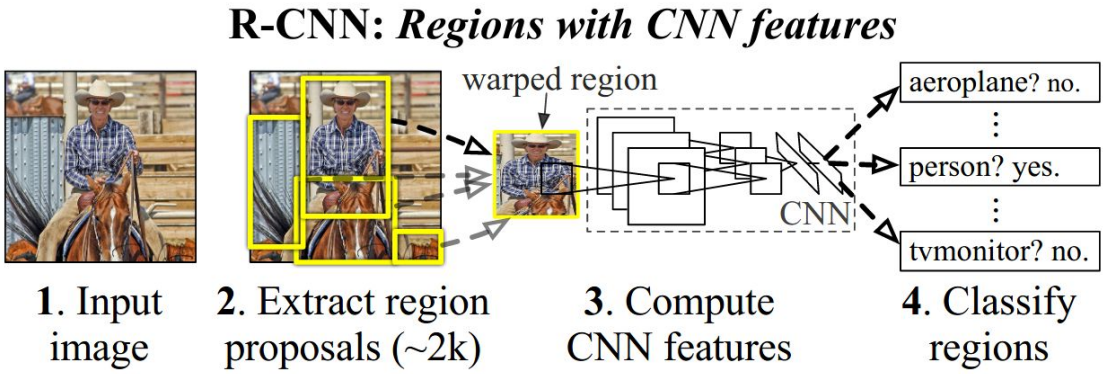
02

Method

Outline of the Method:

The system consists of 3 parts:

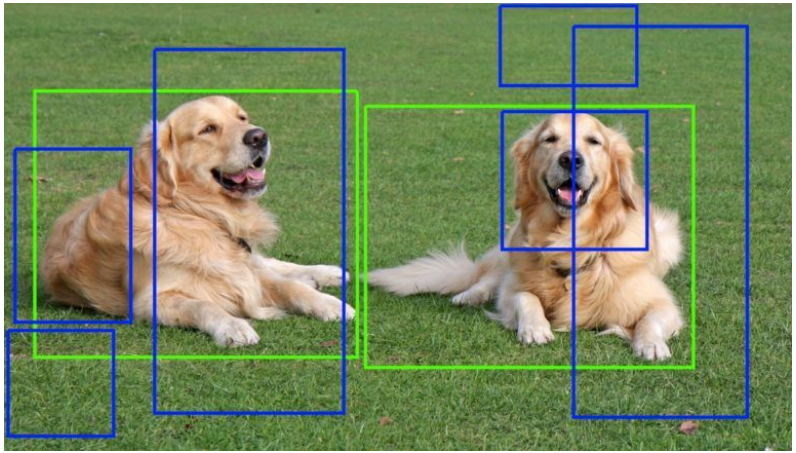
1) Region proposal generation
2) CNN-based feature extraction per region proposal
3) Object Classification



1) Region proposal generation

First of all, what is a region proposal?

- A region proposal is a bounding box candidate that *might* contain an object.



Here the blue and green boxes are the region proposals.

Green proposals are the ones that contain an object.

1) Region proposal generation

We first generate ~2000 proposals per image using selective search.

Selective Search works by over-segmenting an image using a superpixel algorithm.

We could also use other methods like CPMC, exhaustive search, etc.

2) CNN Features

Then for each region proposal we use a pretrained Convolutional Neural Network (CNN) to extract features.

We use AlexNet by Krizhevsky et al. as the CNN (takes 227x227 RGB images, converts them into 4096-dimensional vectors).

To make the image in proposal fit the input constraint (227x227 RGB image), we append p=16 pixels to all the four sides of the region proposal and then resize them to 227x227.

So for each proposal, we get a feature representation of it in the form of a 4096 dimensional vector.

2) CNN Features

Additions to this:

1) We can use a alternate method like center cropping for making the region fit the input criteria of 227x227.

2) We can try alternate feature extractors like VGG16, ResNet50, GoogleNet, Inception-v3, etc.

3) Classification

For the classification of each region (i.e. 4096-dimensional vector), we train one SVM per available class.

One SVM predicts the confidence score of the region belonging to the assigned class.

There might be multiple bounding boxes for the same object. To filter this out we use Non maximum suppression (NMS). This method simply rejects regions if they overlap strongly with another region that has higher score.

Instead of completely removing the proposal with high IoU and losing out on close distinct proposals, we can use an alternate method called Soft-NMS to reduce the confidence score of proposals with high IoU.

03

Goals

Goals

We are trying to achieve the following goals by using a combination of classical tools from computer vision and deep learning.

- To apply high-capacity convolutional neural networks to bottom-up region proposals in order to localize and segment objects.

- To train large convolutional neural networks (CNNs) when labeled training data is scarce by pre-training the network.

Upon completion, we obtain a simple, and scalable object detection algorithm.

Proposed Project Timeline

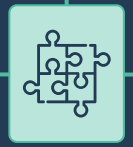
February 20, 2021

Literature Survey, and Related Works

February 28, 2021

Region Proposal Implementation

March 10, 2021



Implementation of Feature Extraction with ConvNets

Proposed Project Timeline

March 28,2021

Implementation of Object Category Classifier(using SVM), and Training

April 10, 2021

Visualization and Analysis of Results

Final Project Presentation

reVision