Spot The Differences Between Two Images

Le Gia Khang
Le Duy Khang
Nguyen Hoang Tan

CS231: Introduction to Computer Vision

December 1, 2023

Presentation Overview

Presentation Overview

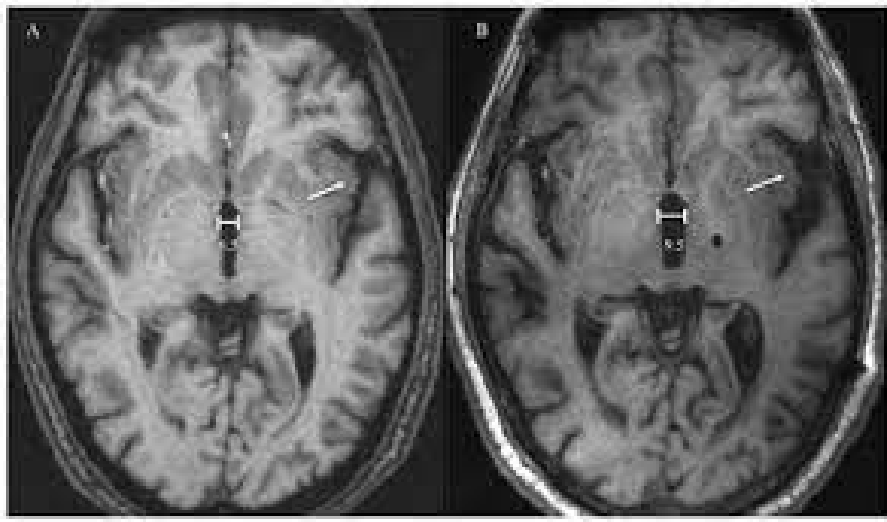
Why it Matters

Applications

Detecting changes is a natural computer vision task:

- The "spot-the-difference" game
- Facility monitoring
- Medical imaging
- Satellite surveillance
- Counterfeit detection

... and many more.



Comparison of two MRI scans over 10-year period.

Input and Output Analysis

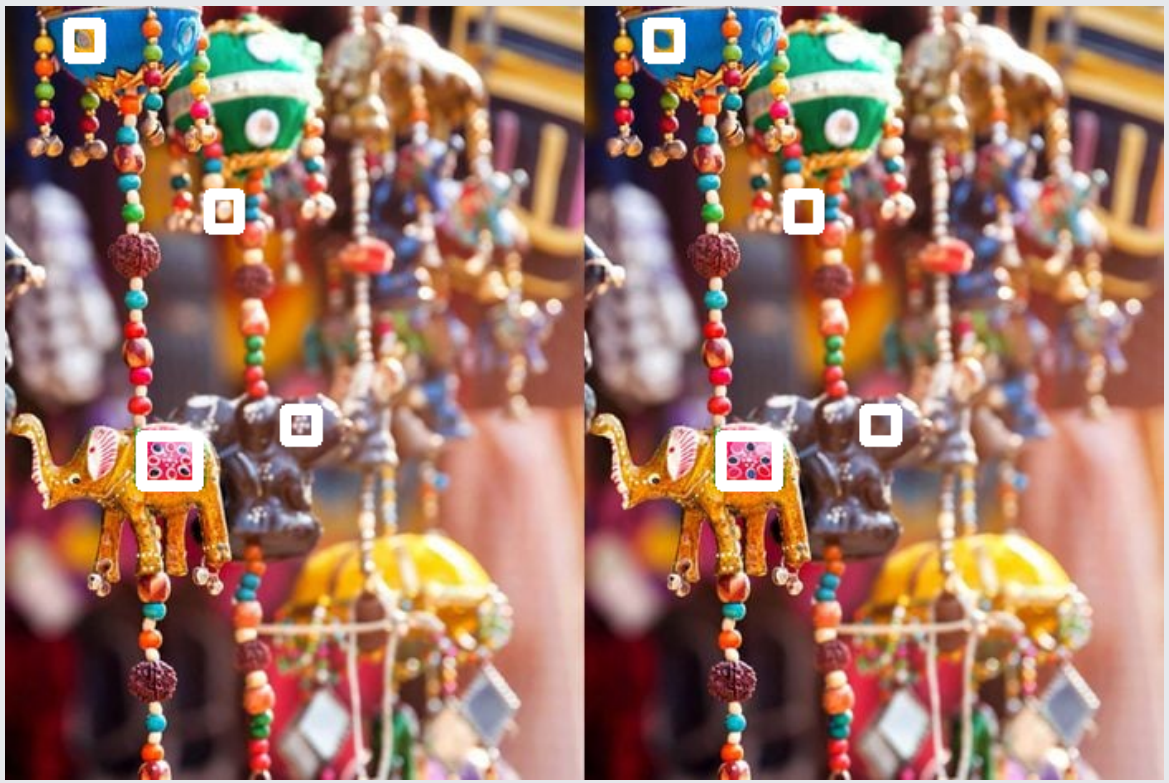
Input

Pair of images need to compare

Input and Output Analysis

Output

The provided images are marked with bounding boxes to indicate areas of distinction

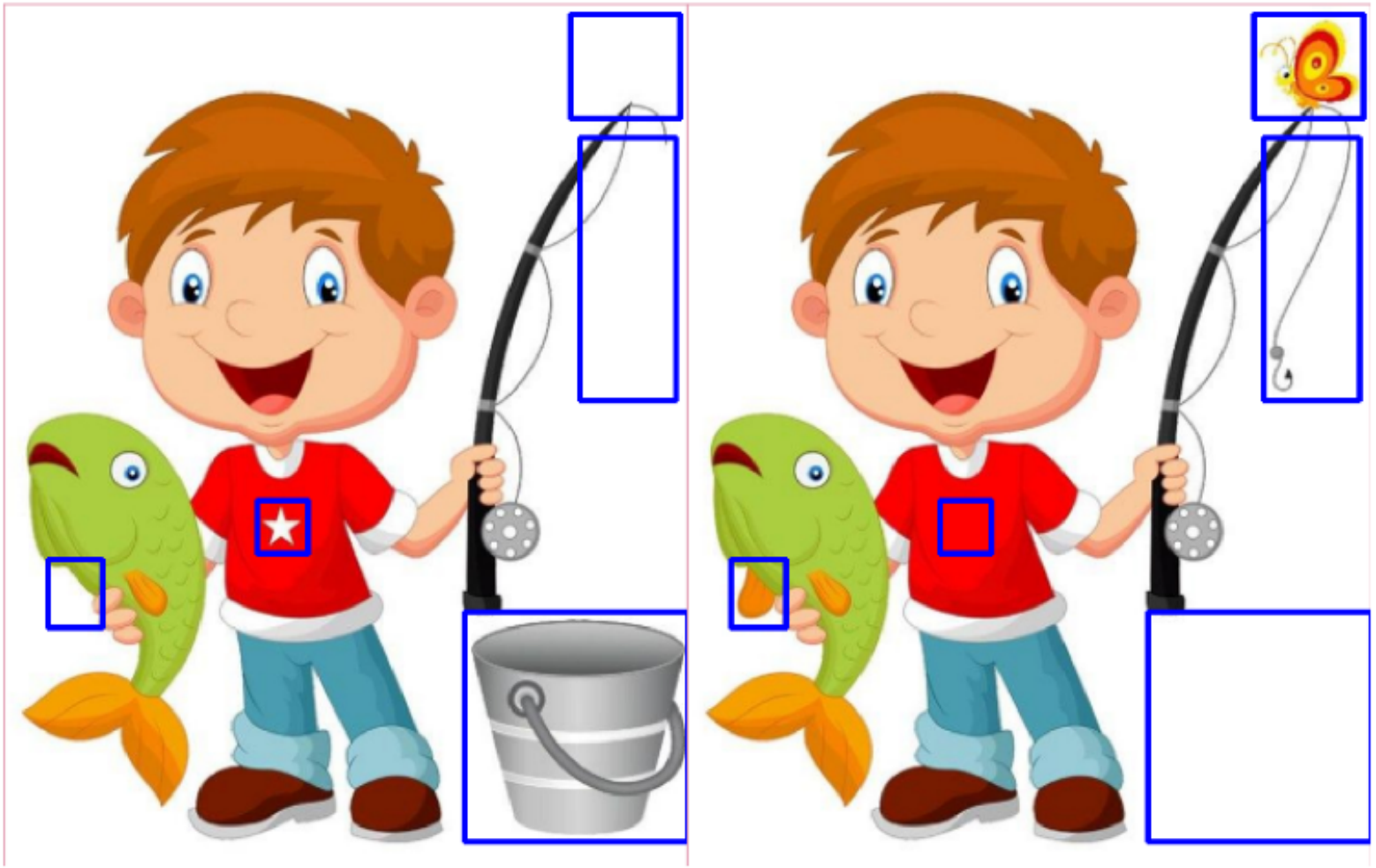


Examples: Input

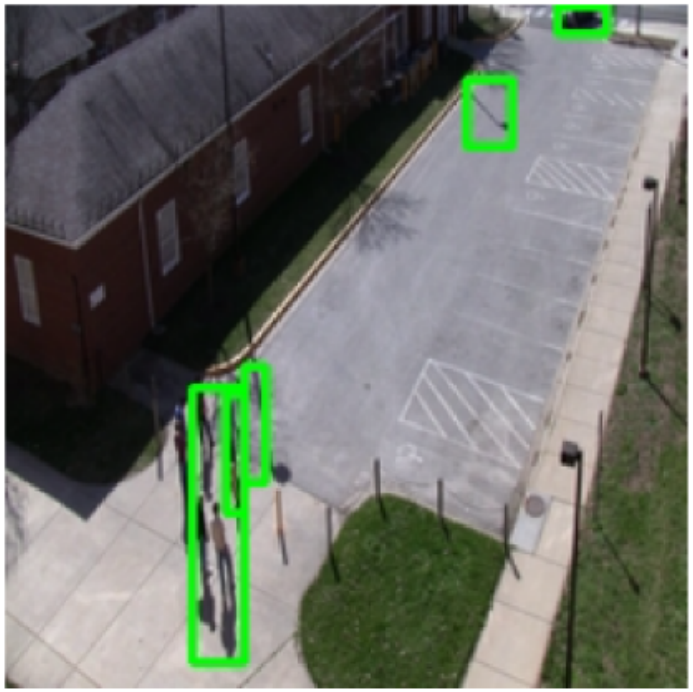
Examples: Input

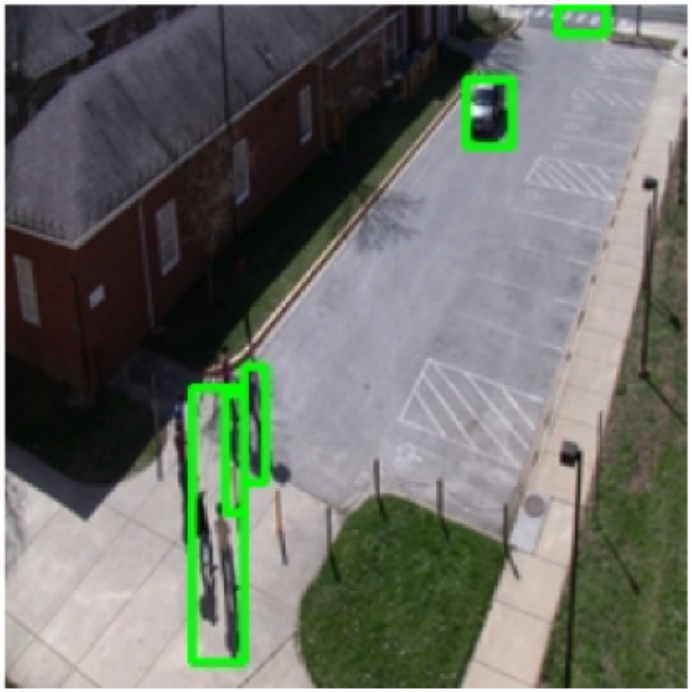
Examples: Input

Examples: Output

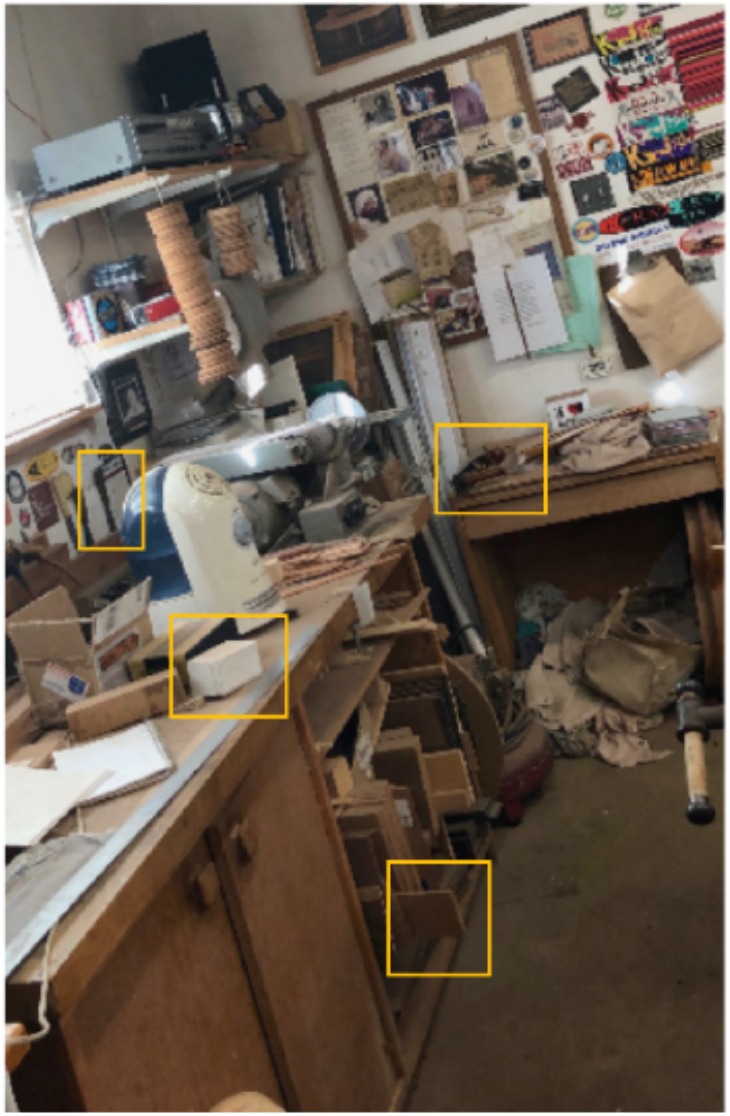


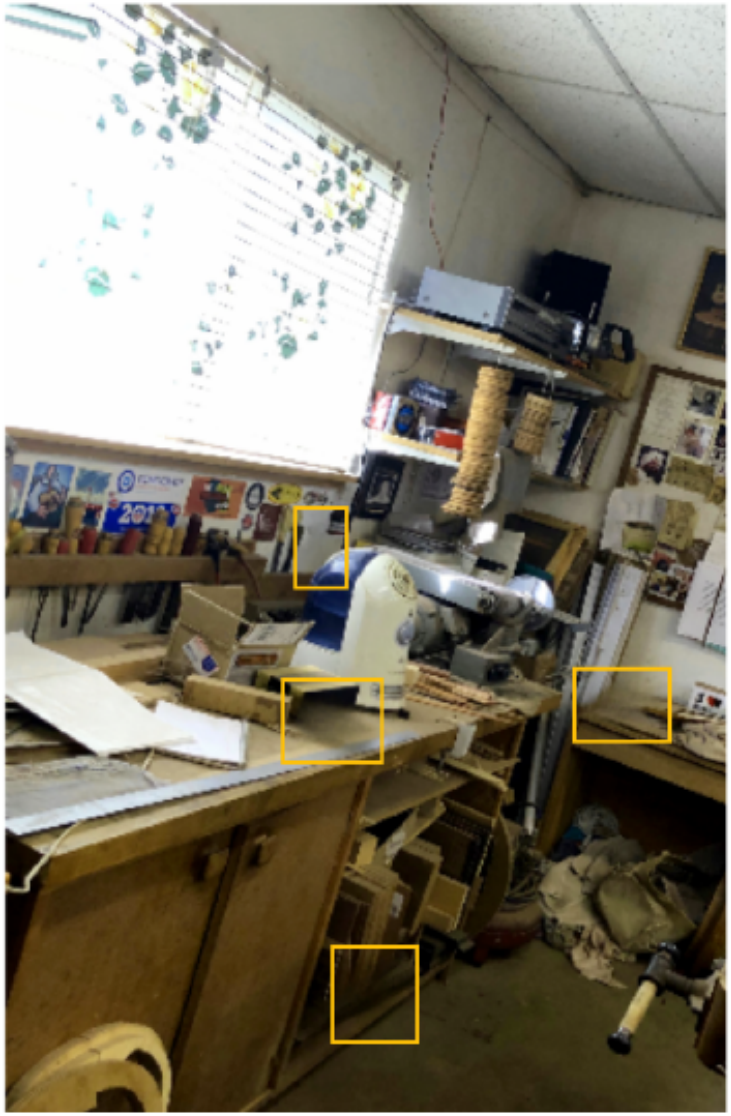
Examples: Output





Examples: Output





Presentation Overview

Siamese Network

Proposed in the article "The Change You Want To See" accepted at the IEEE/CVF Winter Conference on Application of Computer Vision 2023.

Ragav Sachdeva

Andrew Zisserman

Visual Geometry Group at the University of Oxford

Siamese Network: Overview

- Enable "object-level" change prediction and simplify counting the number of changes between two images.
- Use an architecture that operates on two images with geometric (scale, rotation,..) and photometric changes.
- Designed to be class-agnostic, it can detect changes irrespective of the object classes involved.

Model Architecture Overview

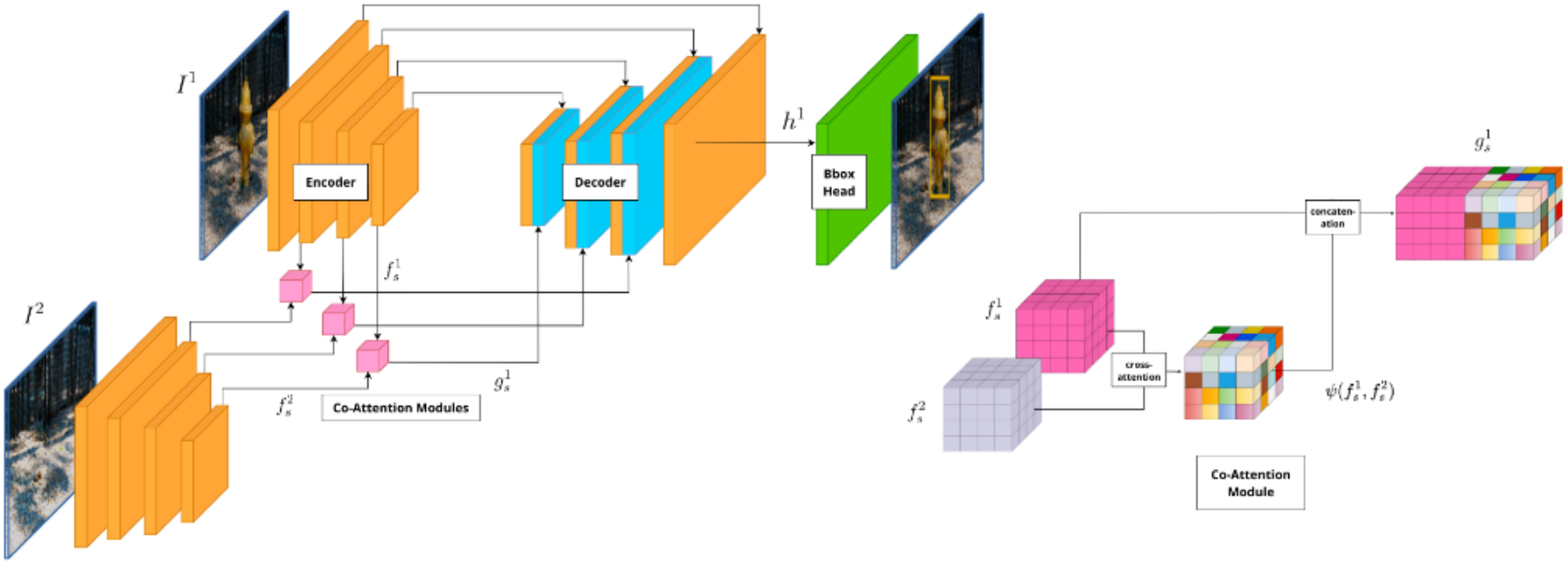


Figure: **Architecture**: Utilizing a dual-image encoder, feature maps (f_1^s, f_2^s) are generated. A co-attention module aligns and conditions these maps (g_1^s, g_2^s). Subsequently, a U-Net decoder processes the original and conditioned maps to yield final feature maps (h_1, h_2). The bounding box detector head employs h_1 and h_2 to generate bounding boxes for images I_1 and I_2, respectively

Siamese Network: Overview

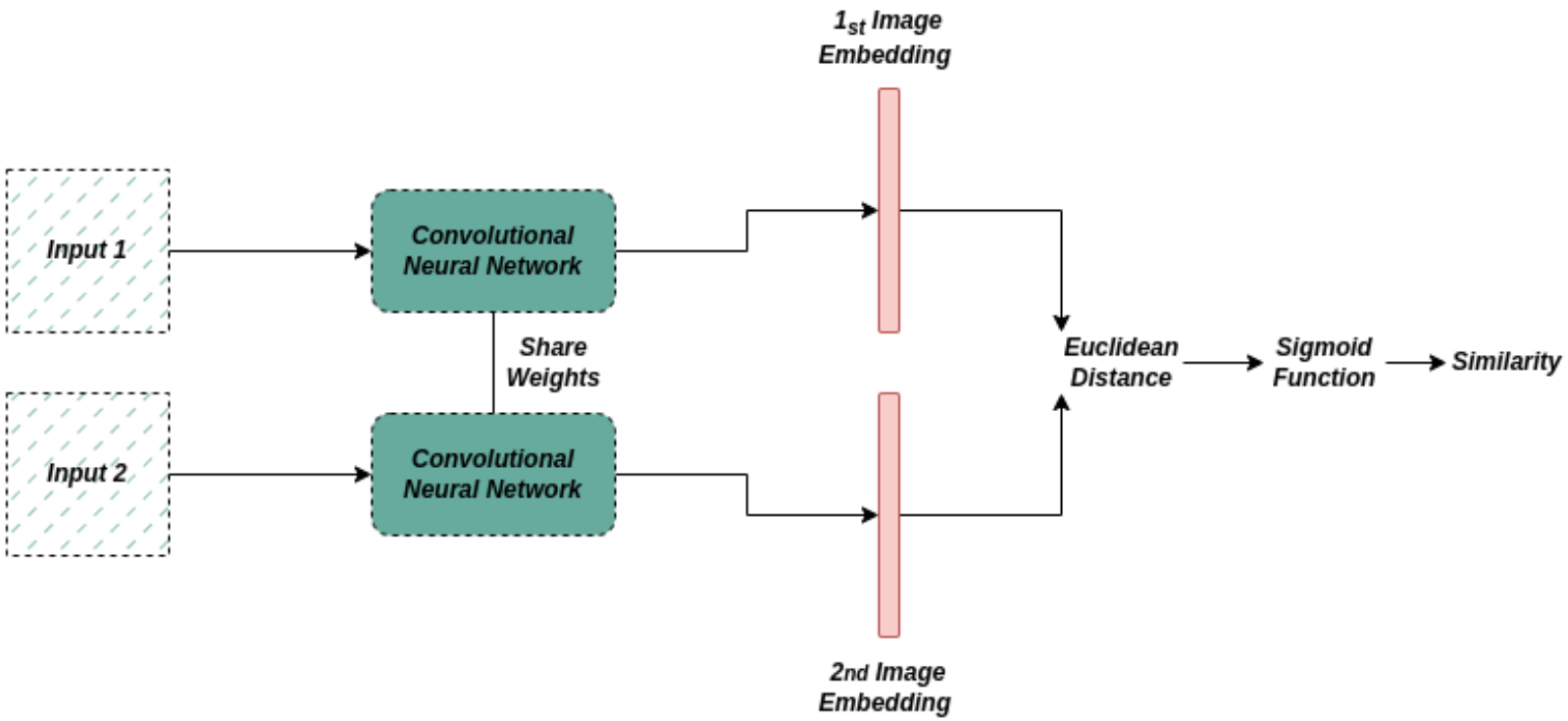
Siamese Network

A type of neural network architecture designed for tasks involving similarity or distance measurement between input pairs.

- Consists of two identical subnetworks (or twins) that share the same set of weights and parameters.
- The name originates from Chang (left) and Eng Bunker (right).

Siamese Network: Architecture

- Consists of two identical subnetworks.
- Extract feature vectors from both networks using a common set of convolutional and fully connected layers.
- Feature vectors from both networks are compared using a loss function L.

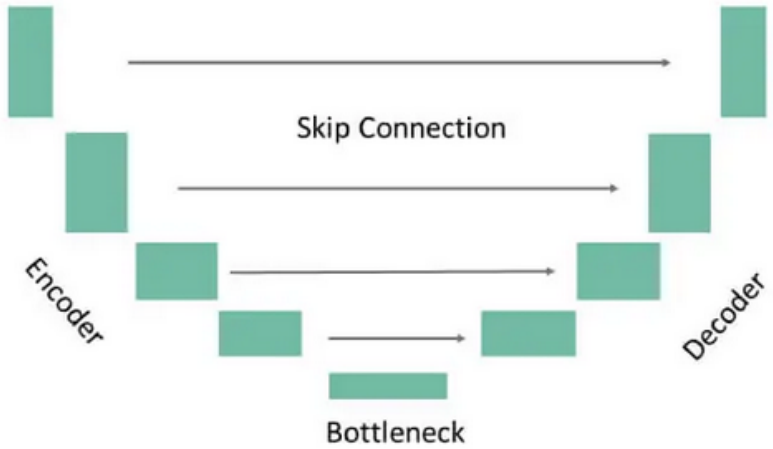


U-Net Encoder-Decoder Network

U-Net is a type of convolutional neural network (CNN) architecture commonly used for image segmentation tasks.

Consists of an encoder-decoder structure:

- **Encoder:** capturing features from the input image.
- **Decoder:** upsampling and producing a segmented output.



The authors employed **ResNet50** as the CNN for the encoder.

A UNet encoder-decoder with CoAM

CoAM Attention Module

We wish to concatenate features from both images in order to condition the model on both input images.

- However, for a given spatial location, the relevant feature in the other image may not be at the same spatial location.

As a result, we use an attention mechanism to model long range dependencies.

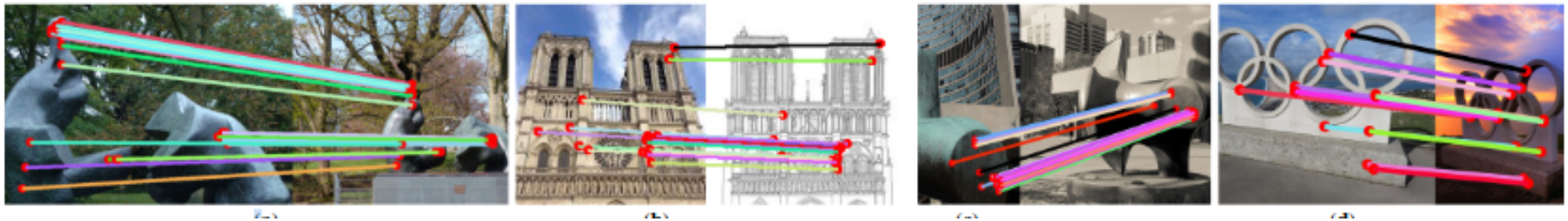
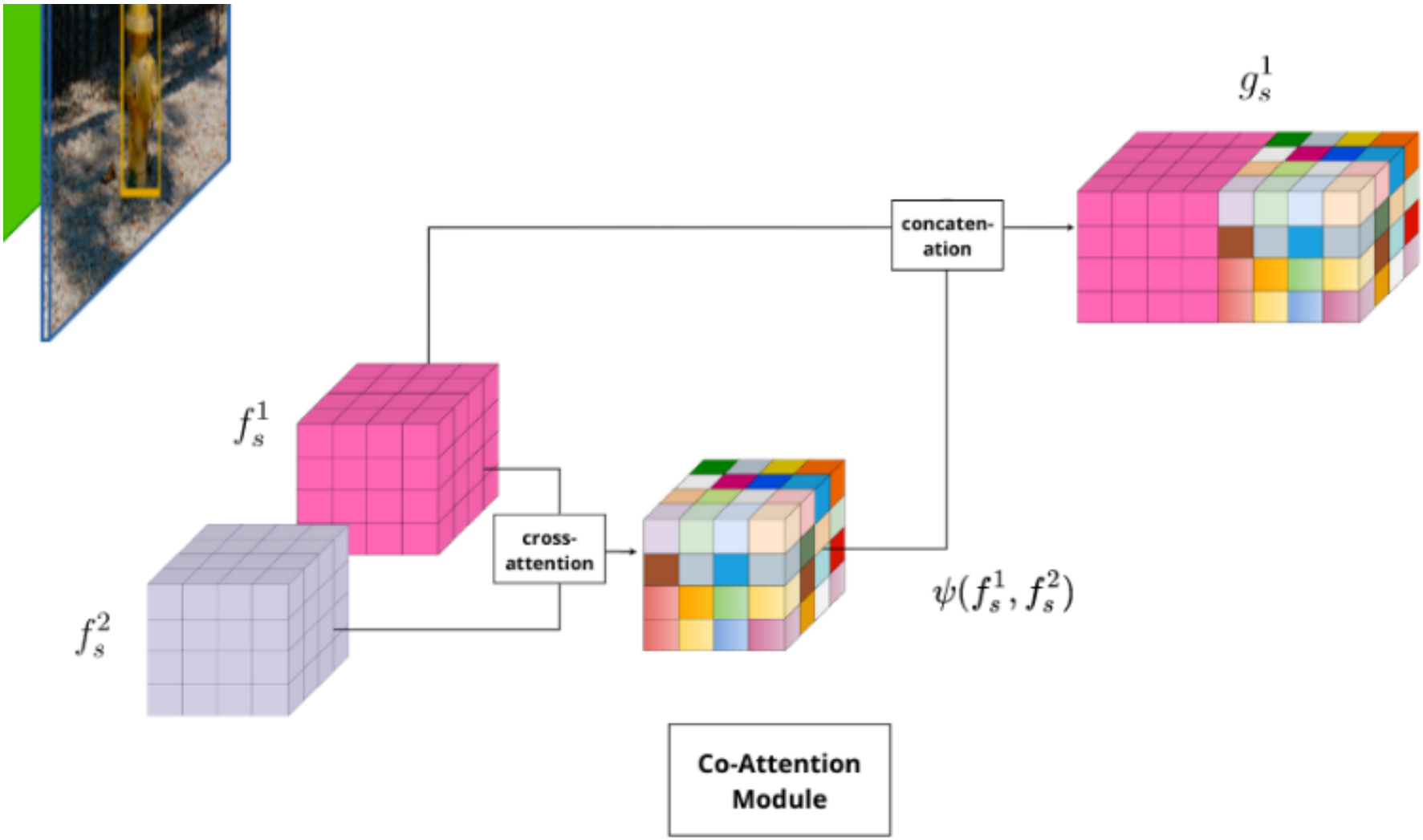
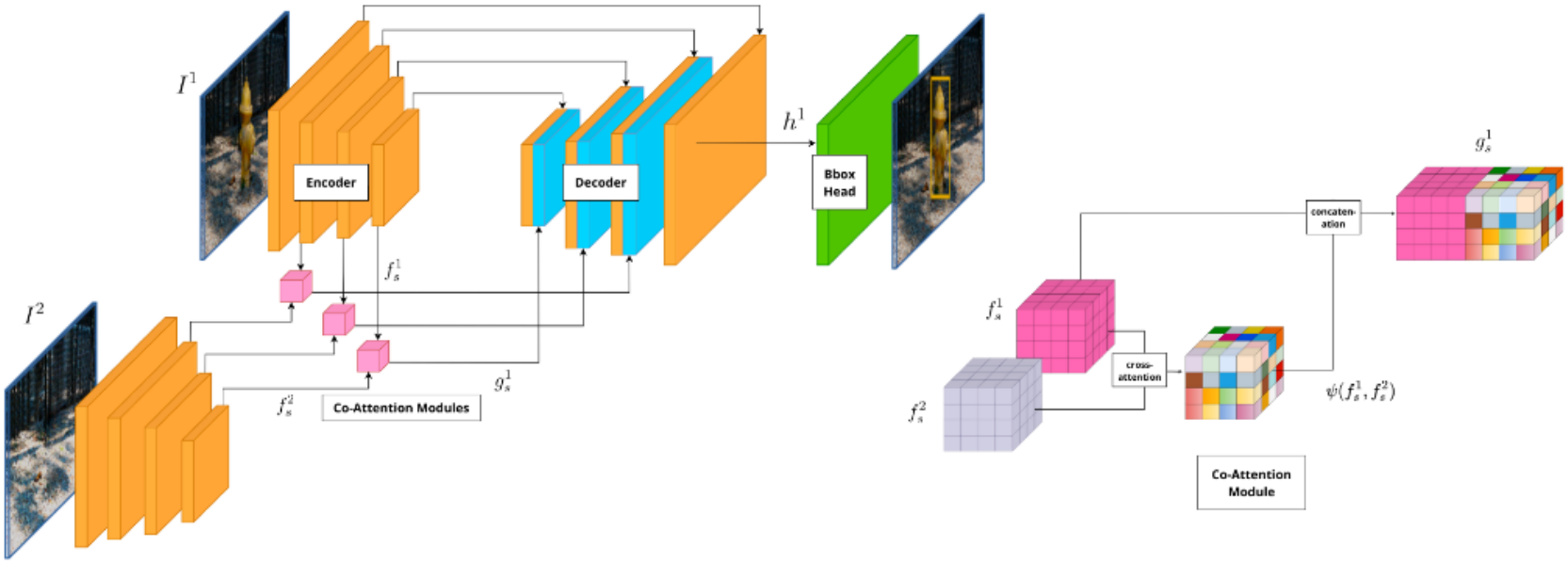


Figure: Correspondences obtained with the CoAM model, which is augmented with an attention mechanism.

A UNet encoder-decoder with CoAM

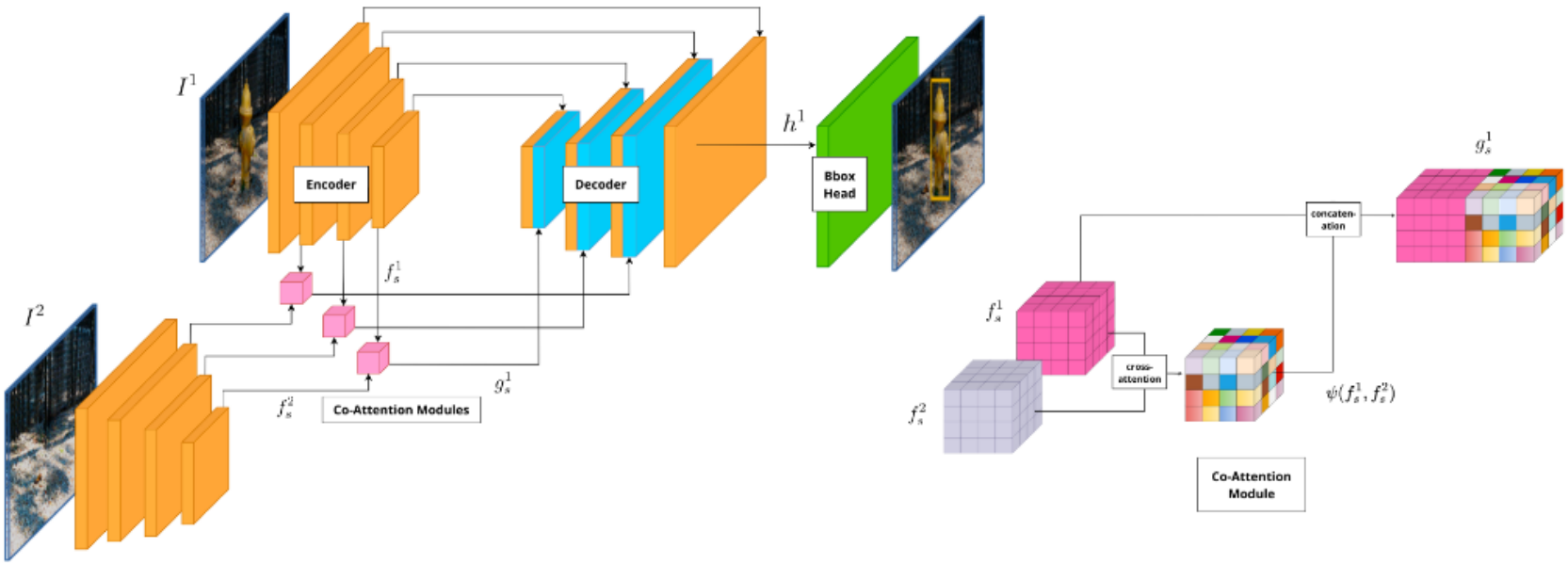


Use Siamese network to detect changes



- First obtaining a set of dense feature descriptors for each image using a CNN-based (ResNet50) encoder.
- These feature are then conditioned on each other using a co-attention mechanism that implicitly supplies the correspondences.

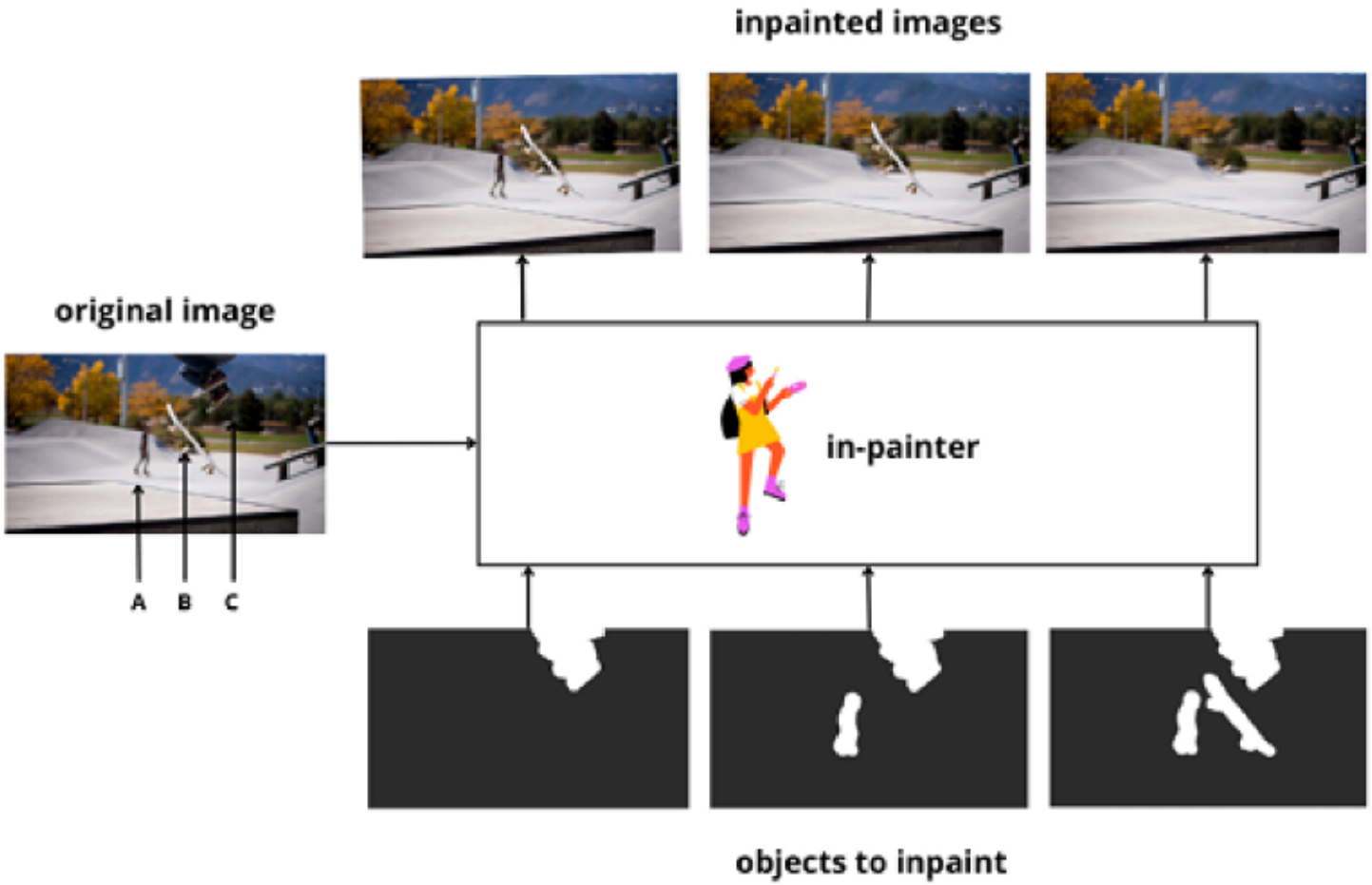
Use Siamese network to detect changes



- Next, feature are passed through a decoder to obtain high resolution conditioned image descriptors which are used by a bounding box detection head to localise the changes.

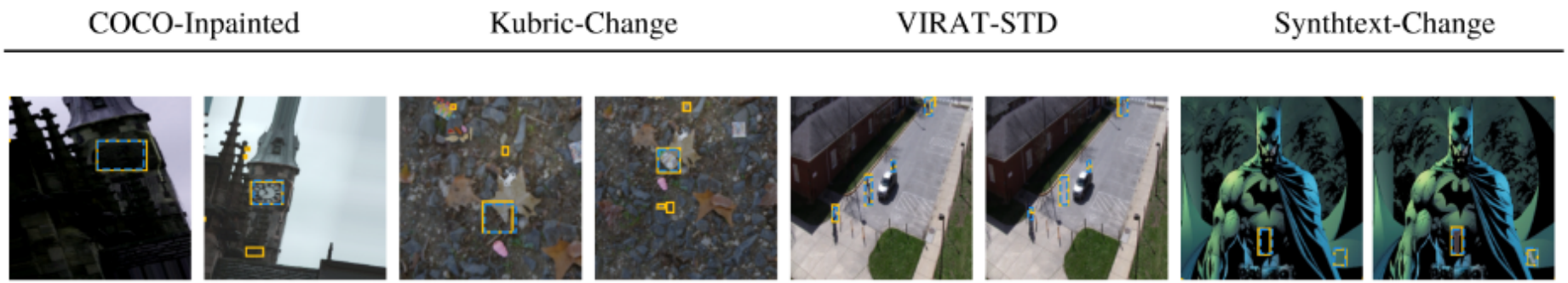
Siamese Network: Dataset

For this method, we make use of a state-of-the-art image inpainting method, **LaMa**, to make the objects *disappear*.



Siamese Network: Dataset

- We also apply random affine transformations to the images along with colour jittering or add random text to "background" images.
- Datasets: COCO-Inpainted, Synthtext-Change, VIRAT-STD, Kubric-Change.



Reference

The Change You Want To See

Ragav Sachdeva and Andrew Zisserman

Understanding SSIM

Jim Nilsson and Tomas Akenine-Möller

Thanks for listening!

Q&A section