Microsoft

Key-point Detection for Measuring Body Size of Giraffes: Enhancing Accuracy and Precision

Solène Debuysere -- 7/26/2023

Contributors: Girmaw Abebe Tadesse, Akram Zaytar, Gilles Quentin Hacheme, Caleb Robinson, Rahul Dodhia, Juan Lavista Ferres

Partners: Derek Lee, Monica Bond

Microsoft | AI FOR GOOD

WILD NATURE INSTITUTE

Microsoft

Introduction & Related works on body size measurements on wild species

AI for Good Lab

Photogrammetry to assess sexual dysmorphism in wild gorillas

Breuer, T., M. M. Robbins, and C. Boesch. 2006. Using photogrammetry and color scoring to assess sexual dimorphism in wild western gorillas (Gorilla gorilla). American Journal of Physical Anthropology

- Photogrammetry is used to measure sexual dimorphism and within-male variation of body length, head size, and sexual dichromatism in a population of wild western gorillas.

Fig. 2. (**A**) Photo showing individual differences of red crest coloration of two silverbacks; silverback to the left has a red crest (score 3) and silverback to the right has a dark red crest (score 2), (photo by Vicki Fishlock). (**B**) Appearance of saddle coloration from different viewing angles and possible scoring of the same silverback photographed under similar light conditions; left: back seated—scoring 3; middle: back quad—scoring 3; right: side—scoring 5. (**C**) Appearance of silverline: from left to right: absence (Silverback OBI), thin (Silverback ZUL), broad (Silverback SAN). (Photos by Thomas Breuer). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

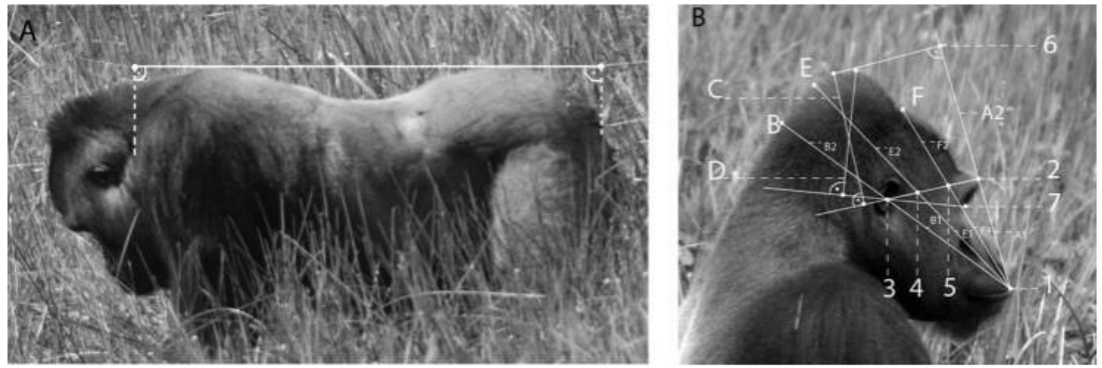
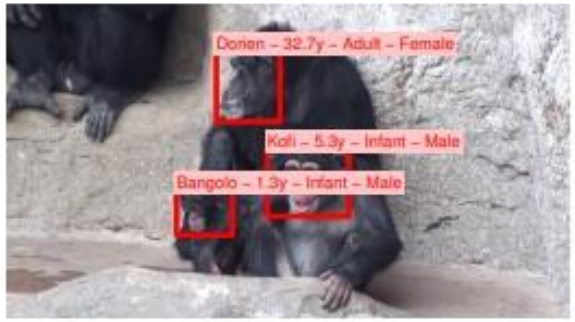


Fig. 1. Photo showing landmarks and linear measurements of (**A**): body length, (**B**): head size. For description of landmarks and measurements see Table 1. (Photos by Thomas Breuer).

Chimpanzee faces in the wild: Log-euclidean CNNs for predicting identities and attributes of primates.

Alexander Freytag, Erik Rodner, Marcel Simon,
Alexander Loos, Hjalmar S Kühl, and Joachim Denzler.
2016. In German Conference on Pattern Recognition.

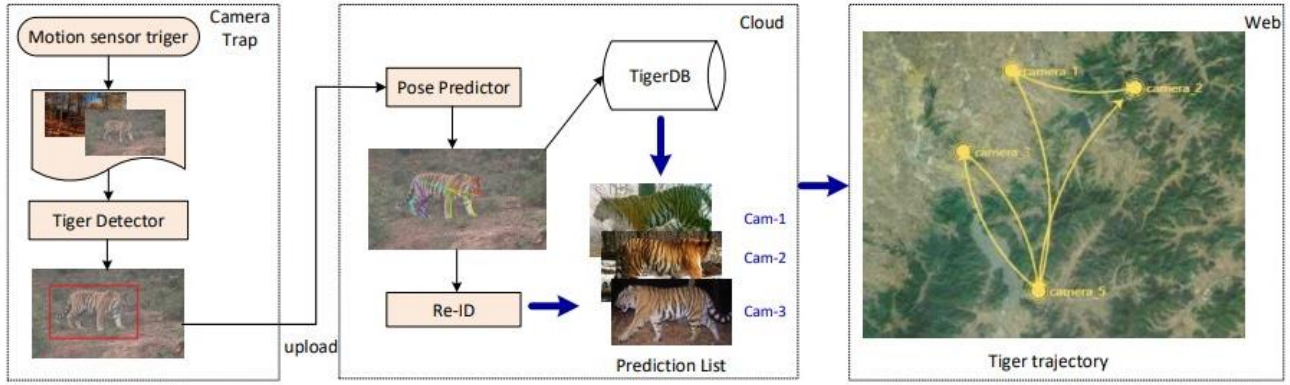
- Computer vision is used to tackled chimpanzee identification, gender prediction, age estimation, and age group classification based on cropped faces. Results have been obtained using ground truth head regions and learned attribute predictions.

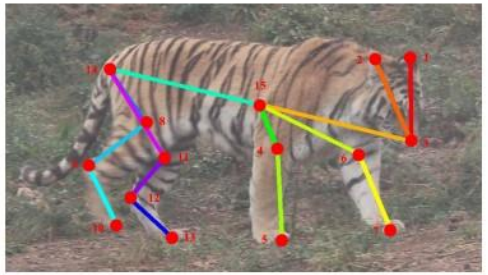


ATRW: A Benchmark for Amur Tiger Re-identification in the Wild

Shuyuan Li, Jianguo Li, Hanlin Tang, Rui Qian, Weiyao Lin

- New large-scale wildlife dataset ATRW, which contains 92 Amur tiger identities from multiple wild zoos, 8,076 high-resolution video clips from multiple cameras in which tigers are annotated with bounding box positions, pose key-points and identities on sampled frames.
- Tiger recognition pipeline, including: Amur tiger detection, pose estimation, and re-identification.

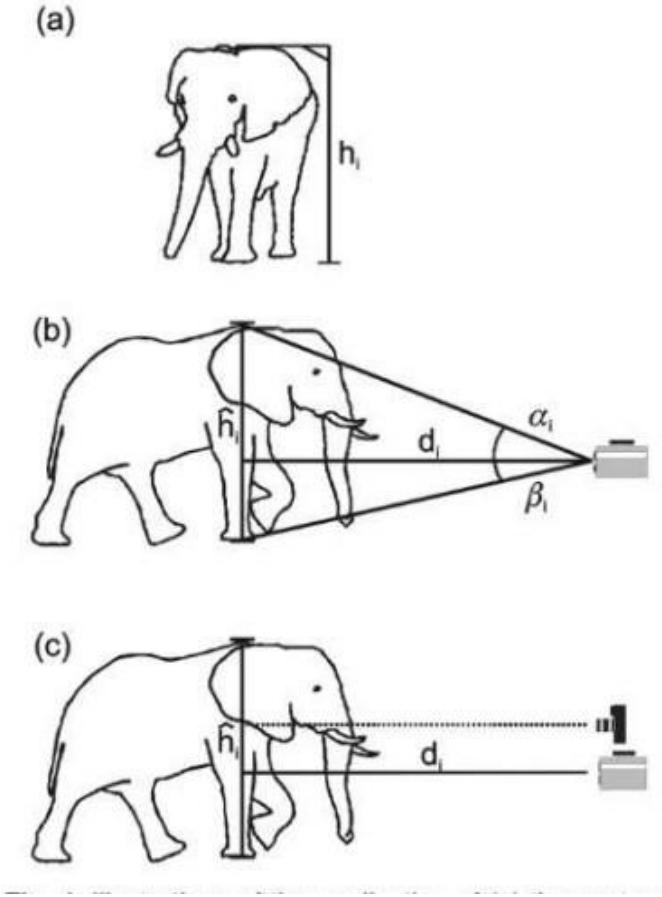




Photogrammetry to measure African Elephants

Shrader, A. M., S. M. Ferreira, and R. J. van Aarde. 2006. Digital photogrammetry and laser rangefinder techniques to measure African elephants. South African Journal of Wildlife Research 36:1-7. Skinner J, Hall-Martin A. 1975. A note on foetal growth and development of the giraffe Giraffa camelopardalis giraffa. Journal of Zoology 177(1):73–79.

- **The custom-built calliper** : consists of a large, sturdy structure positioned next to the elephant with the crossbar resting on the elephant's shoulder.
- **Digital Photogrammetry** : measures the number of pixels that represent the object and uses known distances and focal lengths to calculate the actual size or distance of the object.
- **Laser Rangefinder** : measures the time it takes for the laser beam to bounce back after hitting the object, allowing it to calculate the distance based on the speed of light.



Microsoft

Key-point Detection for Measuring Body Size of Giraffes

AI for Good Lab

Background

Masai giraffes are endangered from illegal killing and habitat loss.

The Wild Nature Institute aims at protecting Masai giraffes through**:**

- **Conservation science:** Conducting scientific studies to understand the behavior, ecology, and population dynamics of Masai giraffes to inform effective conservation strategies.
- **Research**: Investigating the impact of environmental factors, human activities, and conservation interventions on Masai giraffes to develop targeted conservation plans

There are 35,000 Masai giraffes left in the wild today. Their population has fallen by nearly 50 percent in the last 30 years.
PHOTOGRAPH BY SERGIO PITAMITZ, NAT GEO IMAGE COLLECTION

ANIMALS | WILDLIFE WATCH

Masai giraffes declared endangered

The subspecies in Kenya and Tanzania has declined by nearly 50 percent in the last 30 years, largely because of poaching and changes in land use.

BY RACHEL FOBAR

PUBLISHED JULY 12, 2019 • 4 MIN READ

@https://www.nationalgeographic.com/

Problem

Body size measurements of giraffes

- Giraffes have a **unique body form**, making it challenging to collect accurate body size measurements.
- Fundamental questions about **giraffe body size and growth remain** unanswered, such as the shape of the growth curve, asymptotic size, and determinate vs. indeterminate growth.
- Limited availability of **large-scale measurements** on wild giraffes.

Can we measure giraffe proportions, automatically, from imagery?

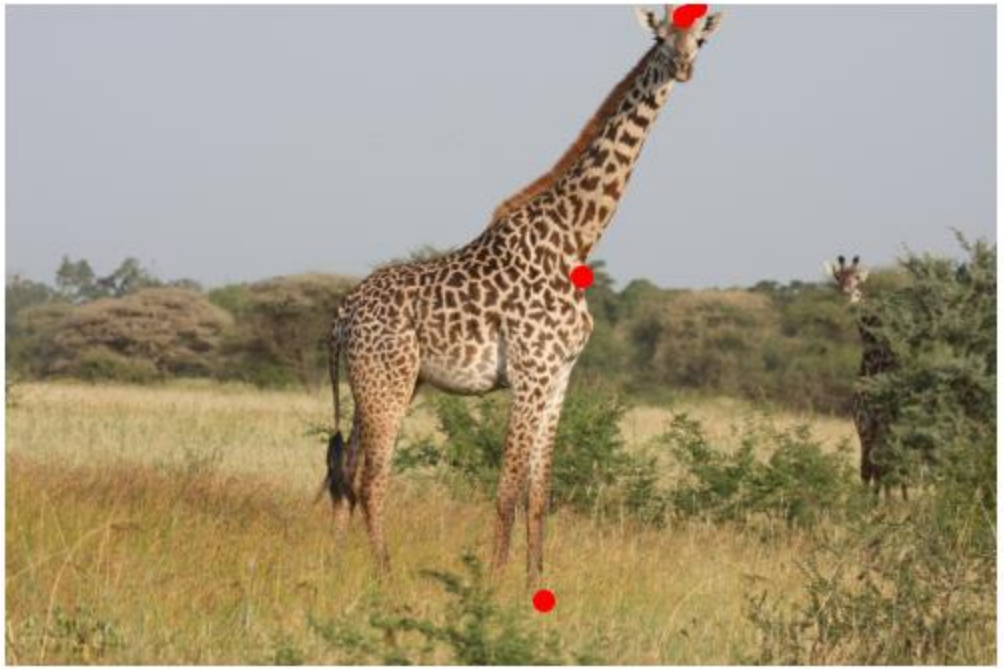
Problem

Body size measurements of giraffes

Load the giraffe image as a tensor of pixels

ML model

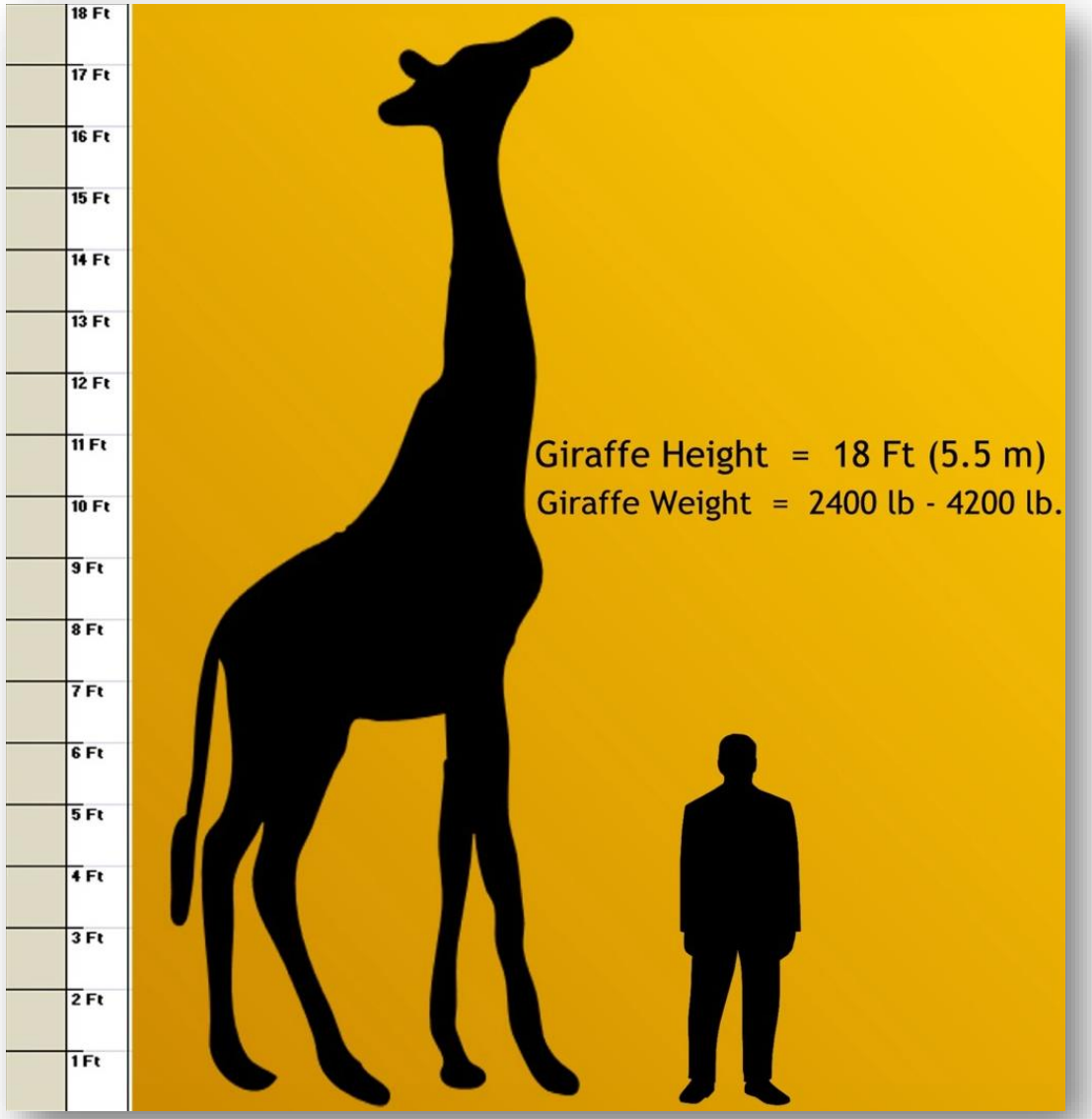
Photogrammetry



Detect 4 key points to measure the height of the giraffe and other values (angles, distances, body proportions, etc.)

Why?

- **Body size measurements** are important for understanding various aspects of **biology**.
- Giraffes are the largest extant ruminant, providing opportunities to study the **effects of extreme body size on mammals.**
- Photogrammetry is an efficient and noninvasive method for measuring giraffe body size.
- Previous studies on giraffe size and growth have relied on small samples, captive animals, or dissected specimens.
- The **need for large-scale measurements** on wild giraffes to improve our understanding of their size and growth.

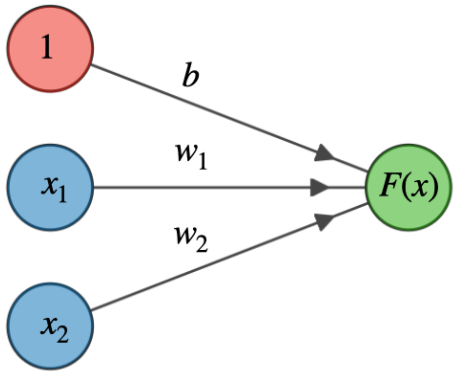


Source: https://www.peoriazoo.org/weights-measurements/

Approaches for key-points detection

USING REGRESSION MODELS

Predict the (x, y) coordinates of each keypoint directly from the image. Examples : MLP, CNNs.

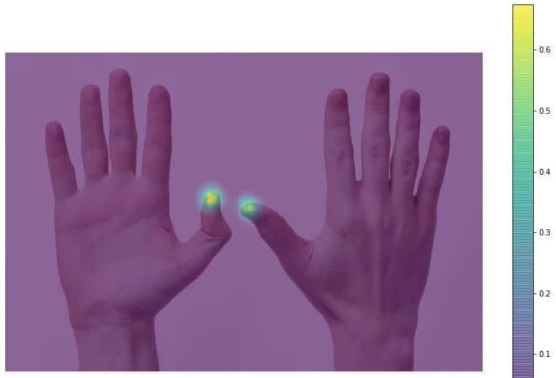


$$F(x) = w^T x + b$$

USING PART AFFINITY FIELDS

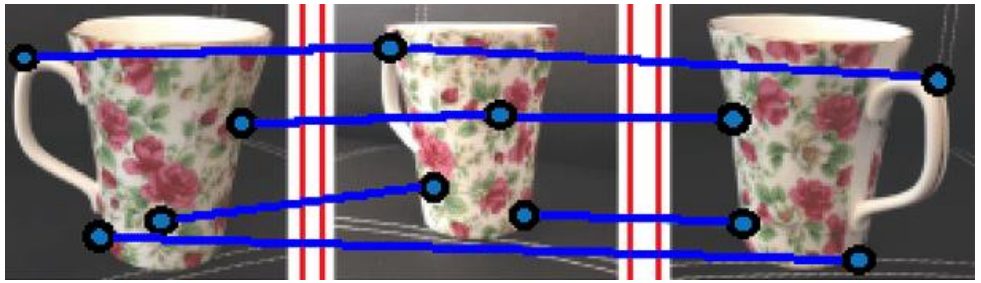
Predict the pairwise affinity between key-points, which describes the likelihood that two key-points belong to the same object. The affinities are then used to group the key-points to estimate the positions Examples : OpenPose (human pose estimation).

Approaches for key-points detection



USING HEATMAP REGRESSION

A heatmap is generated for each key-point with higher values indicating higher probability of the key-point being located at that position. Example : CPMs.



USING FEATURE MATCHING

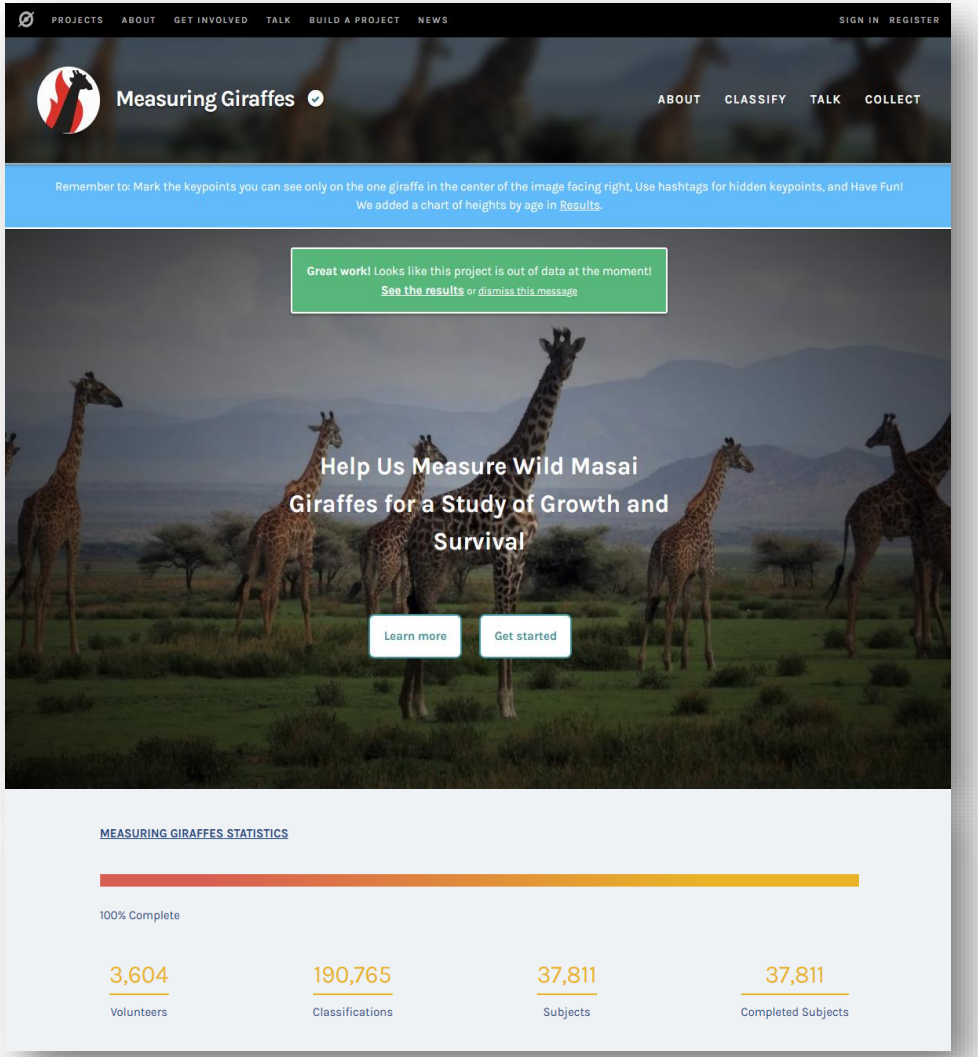
Features such as SIFT or SURF are extracted from the image and then matched to key-points in other images based on similarity. Examples : OpenCV.

Dataset

Source: Crowd-source annotation campaign hosted on the Zooniverse platform (rehosted on LILA science)

Dataset : **WNI Giraffes**

- 29,806 images of giraffes
- 4 key points
  - Front bottom hoof
  - Top ossicone
  - Top head
  - Neck
- ~5 human annotations per key points
  - In reality, it is highly variable



Source: https://www.zooniverse.org/

Overview of the project

1. Exploratory Data Analysis

- Make some statistics on the Dataset
- Define the preprocessing steps

2. Object Detection Model

- Predict bounding boxes coordinates and score, mask, score label
- Link annotations to each giraffe object for each image

3. Keypoints Detection Model

- Predict 2 coordinates values for each of the 4 keypoints

4. Deployment of the Model

- Deploy Gradio Demo on Hugging Face
- Create Executable python file Script for the partners

Challenges

Dataset

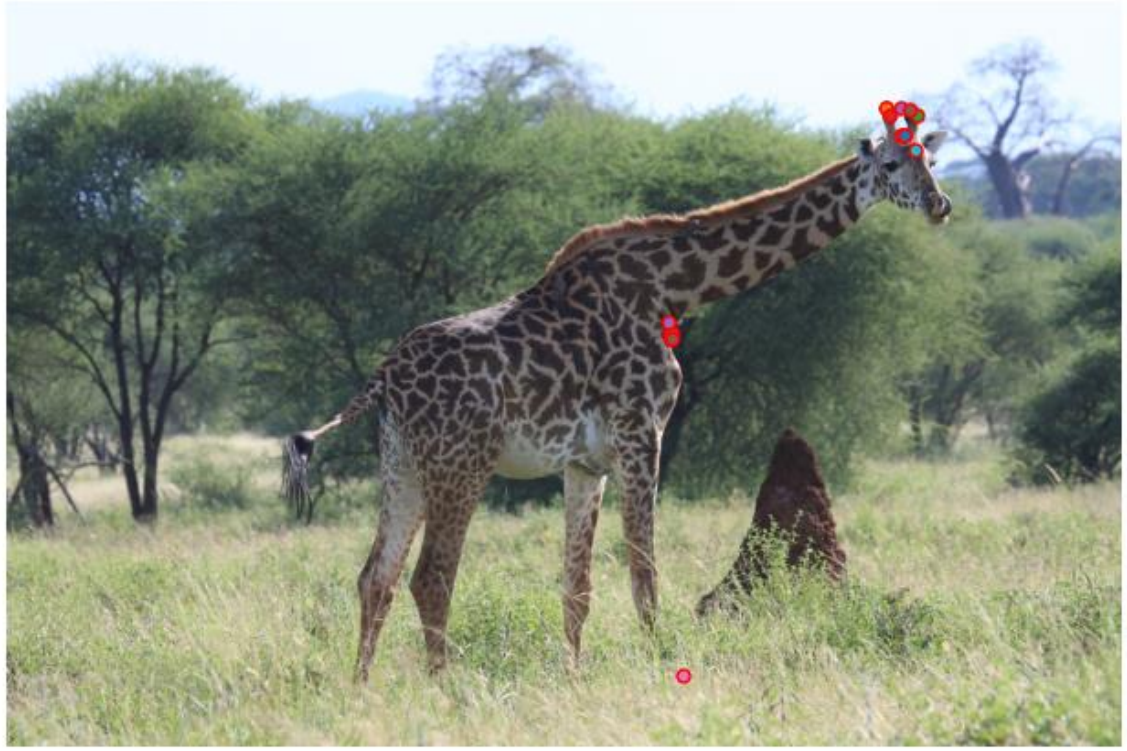
Outliers sizes

29 806 RGB images

- 53 % images with size (5472, 3648) : group A
- 28 % images with size (5184, 3456) : group B
- 18 % images with size (3888, 2592) : group C
- 0.3 % images with size (4896, 3672) : group D
- 0.003% images with size (2592, 3888) : group E

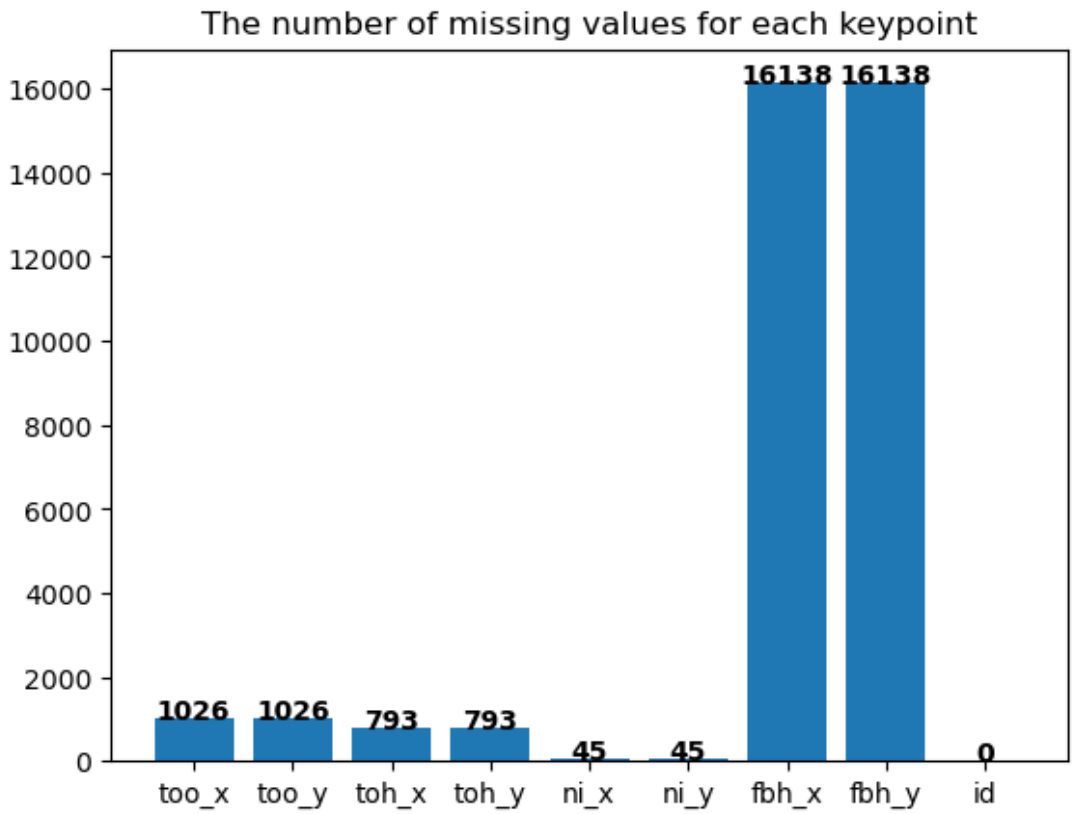
Preprocessing step :

Keep only the 99% images with the ratio Height/Width = 1,5



Challenges

Dataset



Keypoint's annotations

4 keypoints (x, y) :

- Too : top of ossicones
- Toh : top of head
- Ni : neck indent
- Fbh : front bottom hoof

Original description of the Dataset :

- 5 persons annotated each keypoints
- 1 Giraffe only must be annotated on each image

Around 55% of images have at least 1 missing keypoints

Data challenges

Challenges on images

- Images have **really high dimensions** and **different sizes** that may lead to distortions when resizing
- Imbalanced dataset with giraffes that have **different orientations** (vertical, horizontal right or left)
- Imbalanced dataset with giraffes of **different sizes** (giraffes really far in the background or really close to the camera)
- Some **giraffes are cut** and the **annotations are random** in the vegetation
- Imbalanced dataset with images with **more or less contrast** or **low visibility** that makes the giraffe more difficult to identify in the background
- Some images contains **more than one giraffe** and they are crossing or overlapping each other

Challenges on key points

- Some giraffes are **not correctly annotated** :
  - Confusions between left and right (left top ossicones, right front bottom hoof)
  - Random annotations on the vegetation (for example when the head of the giraffe is behind leaves)
  - Some annotations are on the back of the giraffe instead of the front bottom hoof
- For each image, **all annotations are mixed together** in the JSON file even if there are more than one giraffe
- 55% of samples have at least one **missing keypoints** (total number of annotations equals to zero)
- The **total number of annotations** per keypoints is **random** (not 5 annotations per keypoints as the Dataset is described)
- 28 images have **30-50 random keypoints** on it

Challenges

Preprocessing steps

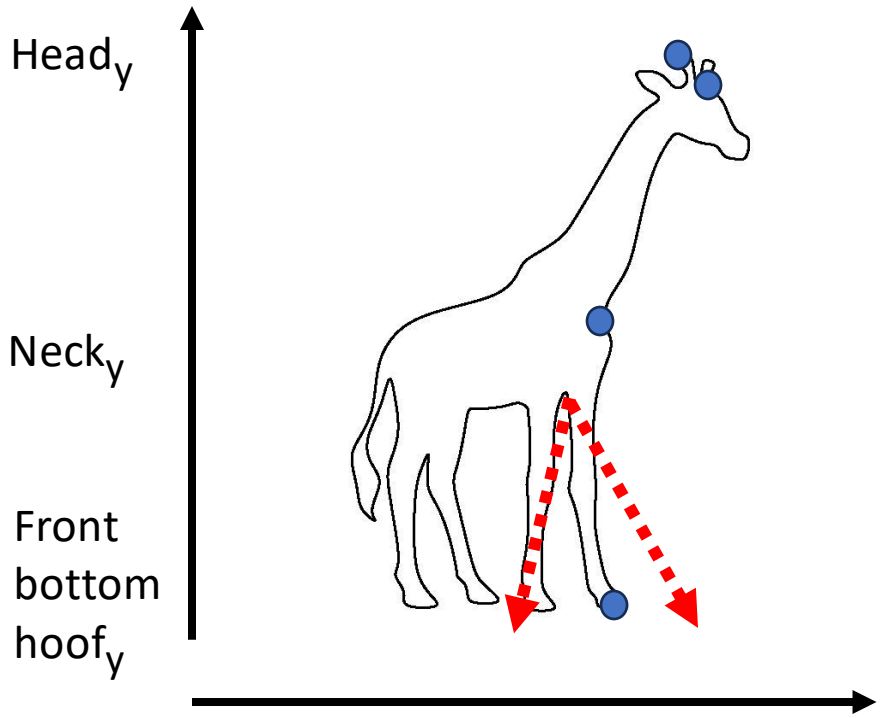
Outliers

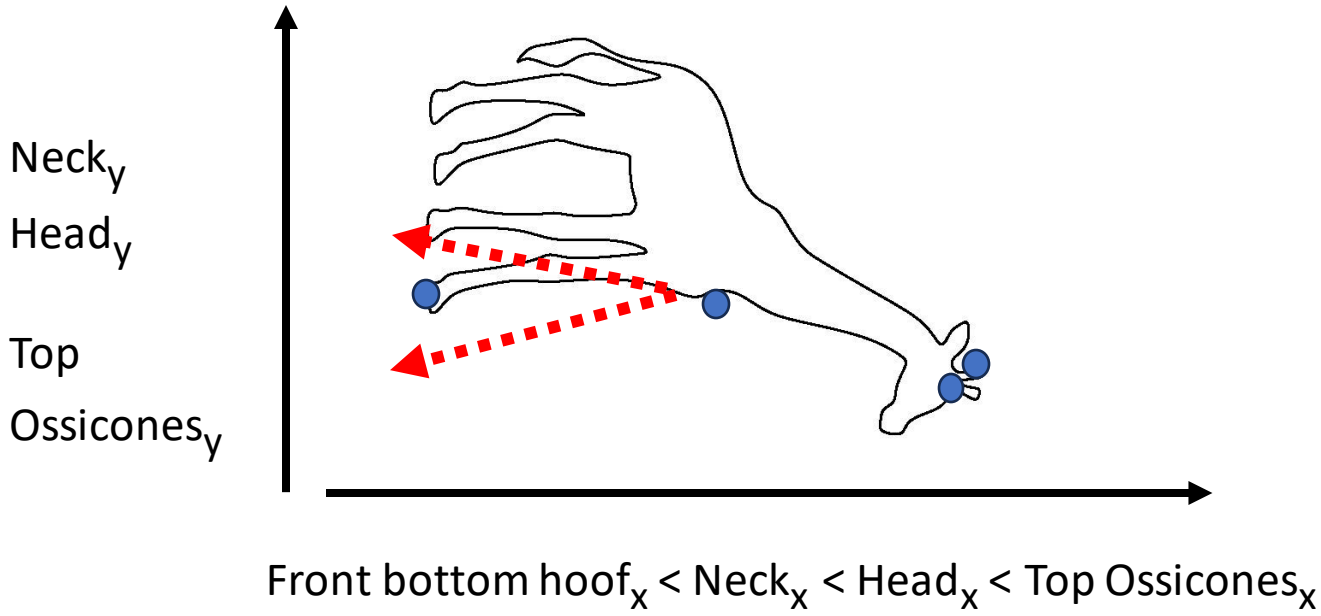
Problem: Some Giraffes are horizontal with different orientations (head on the left or on the right)

Challenges

Preprocessing steps

Giraffes orientations





Front bottom hoof$_x$ < Neck$_x$ < Head$_x$ < Top Ossicones$_x$

(Front bottom hoof$_x$) < Top Ossicones$_x$ < Head$_x$ (< Front bottom hoof$_x$)

Challenges

Preprocessing steps

Outliers

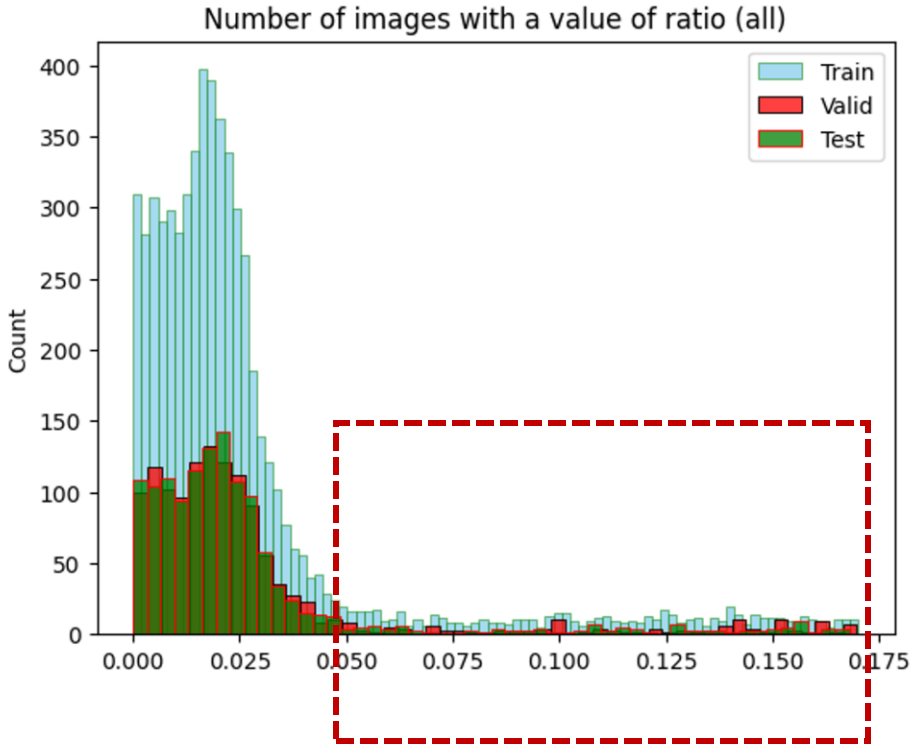
How can we remove giraffes that are cut or rotated?

- Use ratio to detect outliers and horizontal images (2 cases : head on the right or left)

$$\text{delta\_x} = |\text{too\_x} - \text{toh\_x}|$$
$$\text{delta\_y} = |\text{toh\_y} - \text{fbh\_y}|$$
$$\text{ratio} = \frac{\text{delta\_x}}{\text{delta\_y}}$$

- High value of ratio implies a small value for delta_y so the giraffes might be horizontal or cut

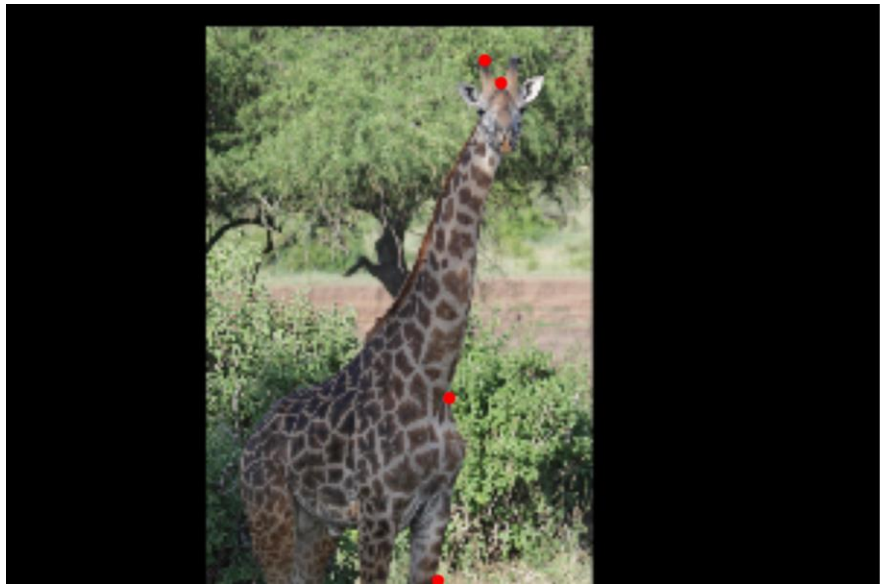
- These samples are the cause of really high MSE during the training of the ResNet model

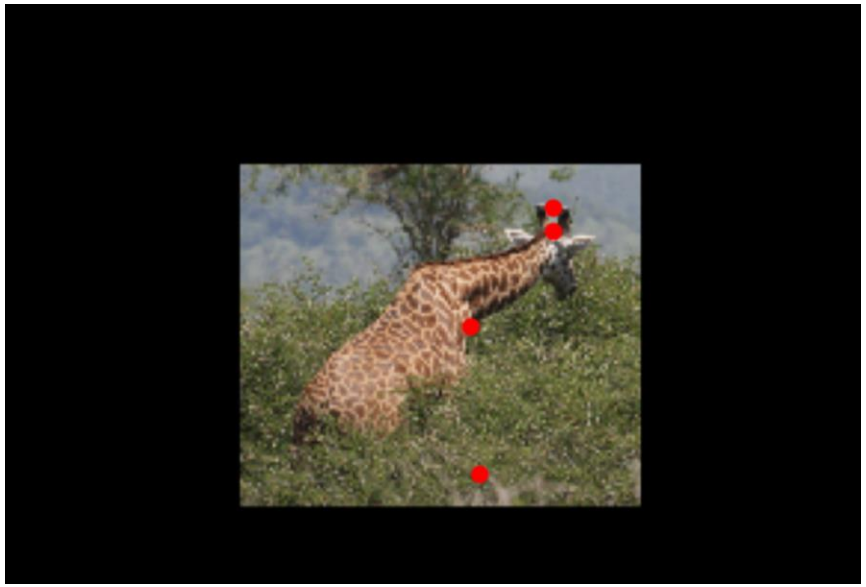


Challenges

Preprocessing steps

Outliers



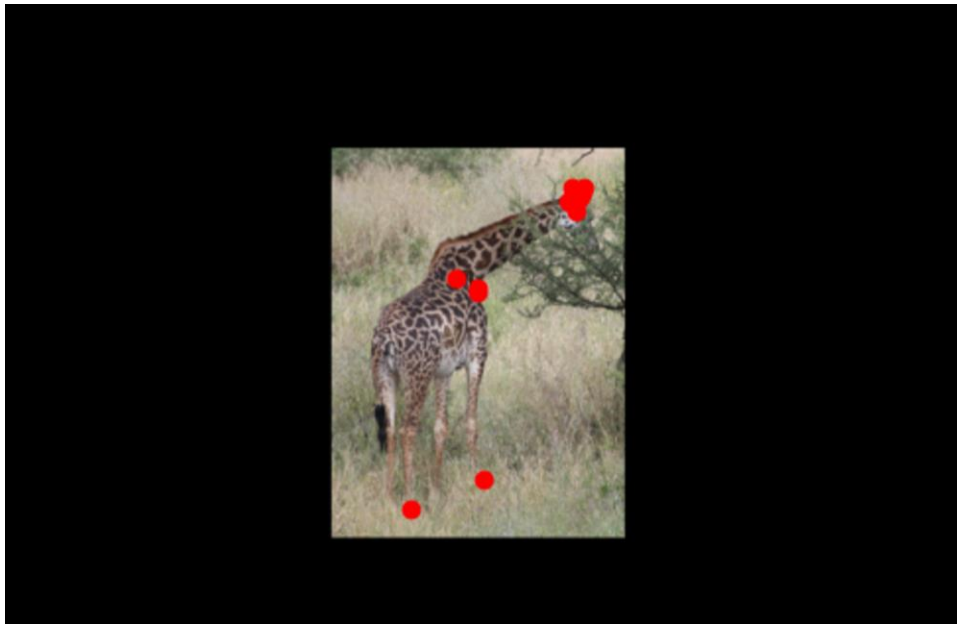


Problem : Some giraffes are cut and the annotations are random

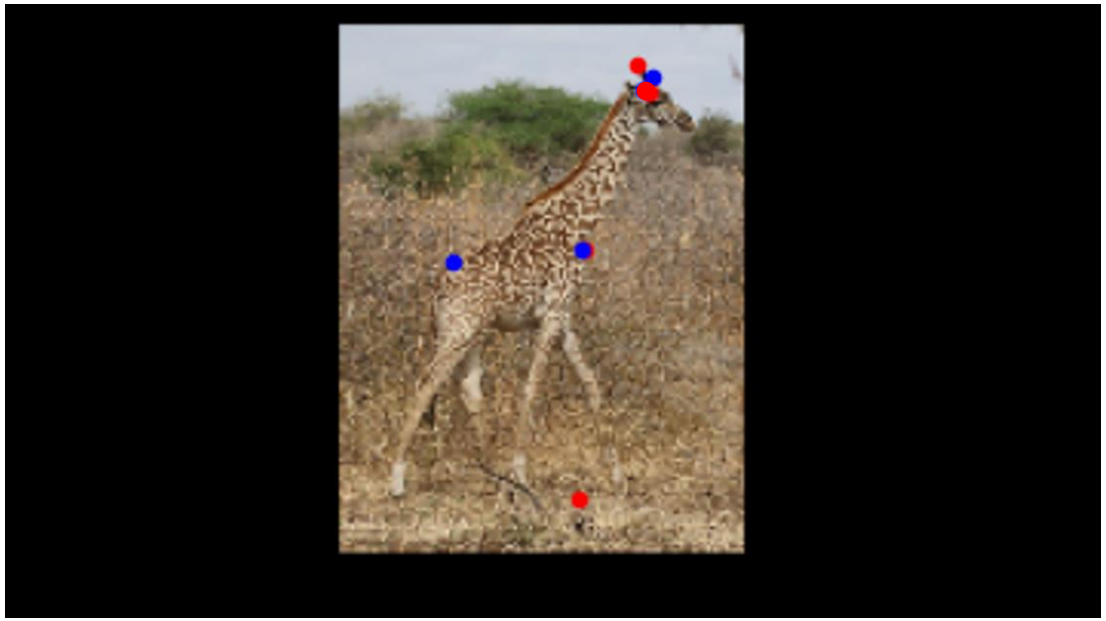
Challenges

Preprocessing steps

Outliers



Problem : The annotations is not on the correct hoof

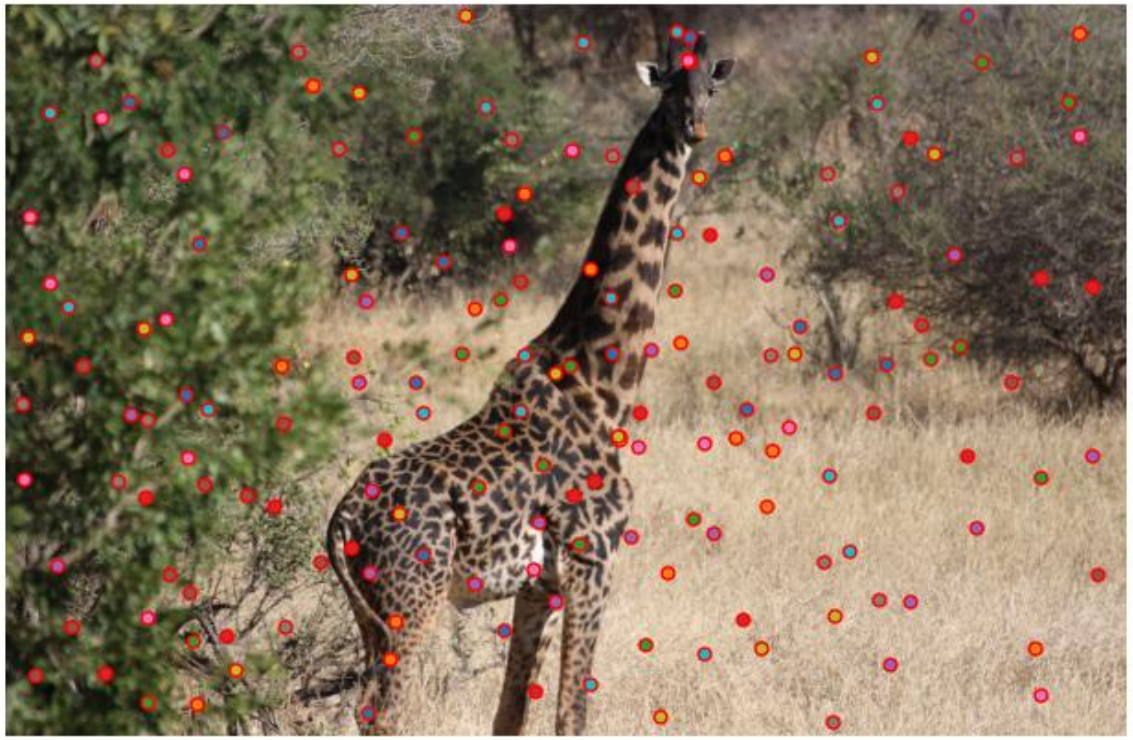


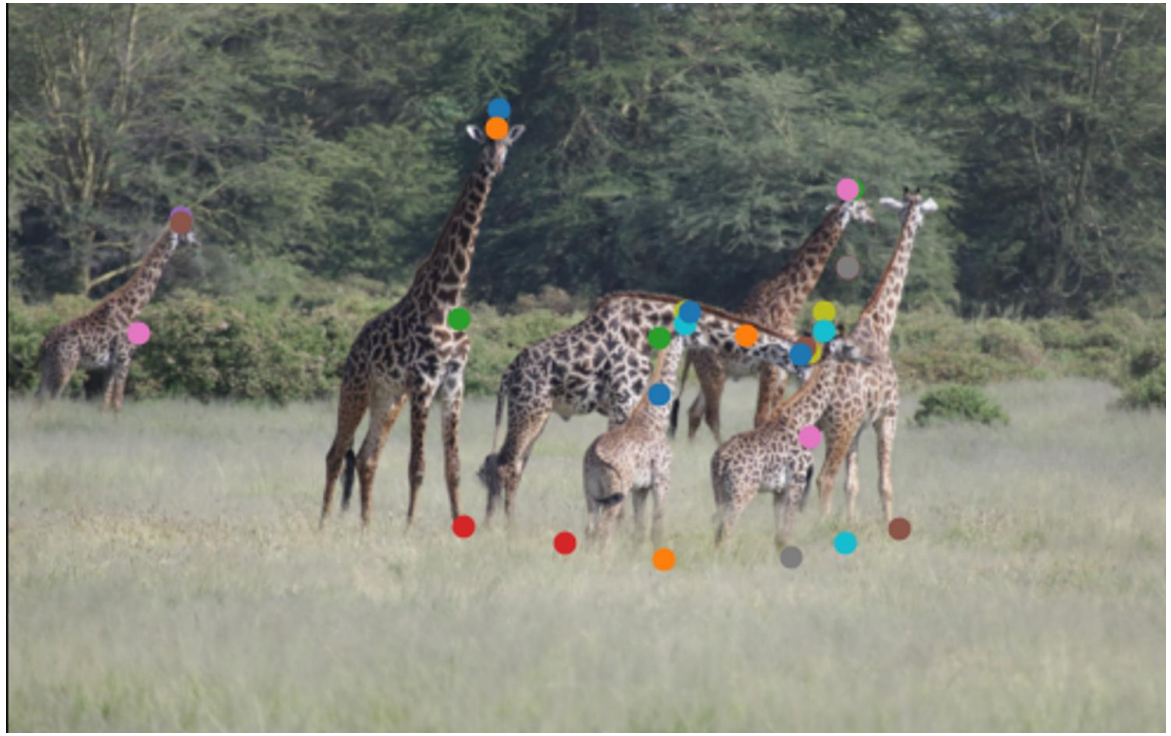
Problem : The annotations is not on the hoof but on the back of the giraffe

Challenges

Dataset

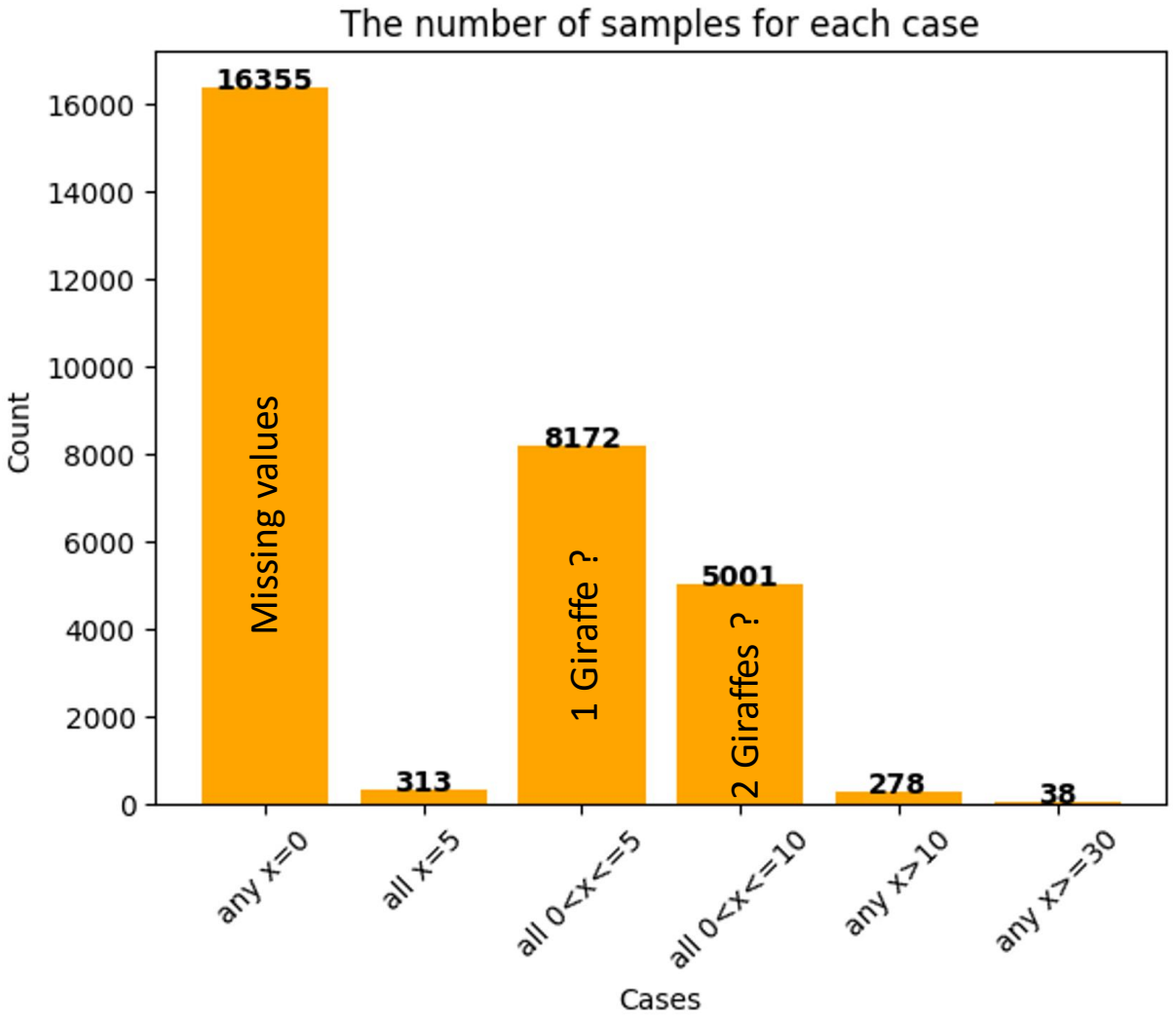
Outliers





Challenges

Dataset



Keypoint's annotations

Observations on the Dataset :

X = total number of annotations

- 1 Giraffe : any sample where 0< X <=5
- 2 Giraffes : any sample where 0< X <=10 and at least one keypoint X > 5
- Outliers : any sample where X>10

Problem :

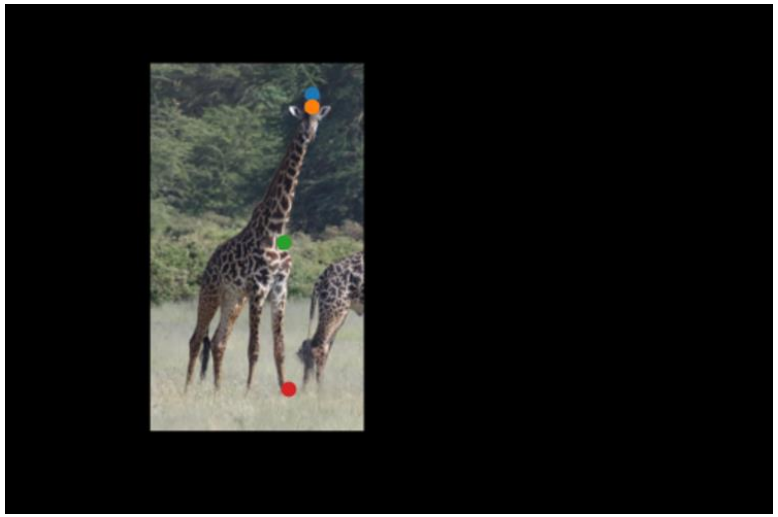
- The annotations of all the giraffes in an image are **mixed**. We can't use the median of the annotations to train the model.
- This statistical approach is too approximative to estimates the number of outliers without removing good samples.
- Around 55% of images have at least 1 missing keypoints

Approach

Object Detection Model

Pretrained MaskRCNN to link annotations to each Giraffe

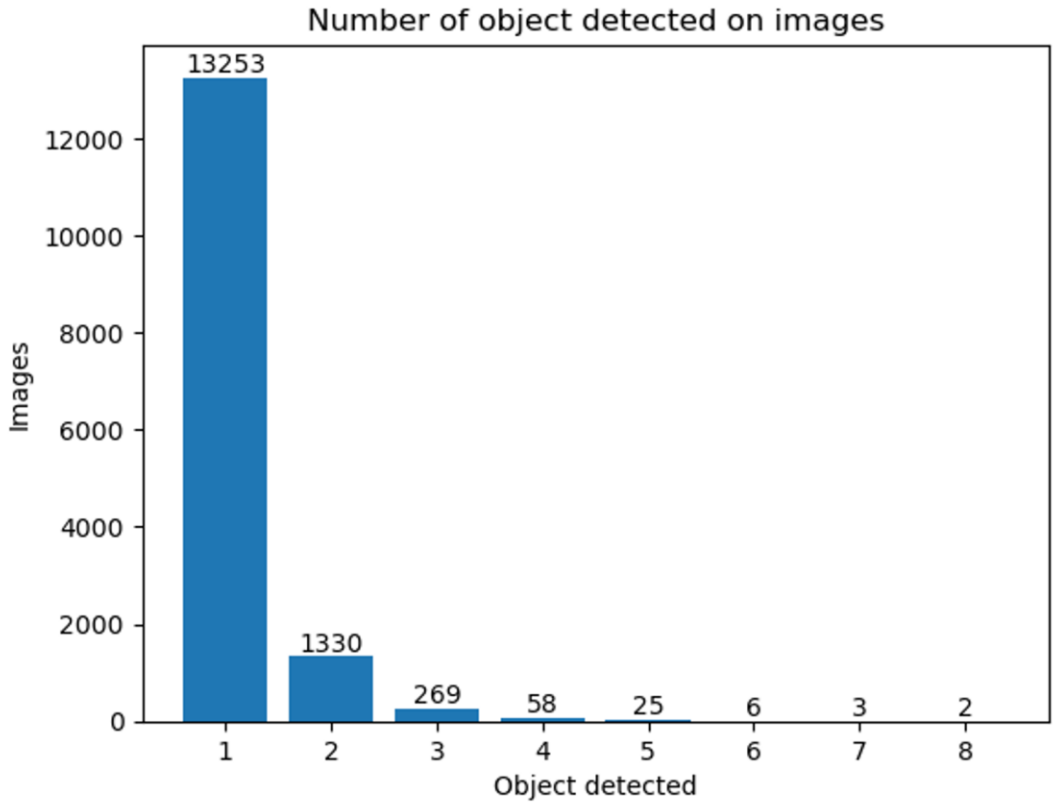
1. Predict the bounding boxes coordinates
   - Mask RCNN pretrained on COCO Dataset
   - Predict bounding boxes coordinates and score, mask, score label
2. Create new samples from one image
   - Crop the image to keep only the giraffe object
   - Associate the keypoints to the correct bounding box
3. Create the final JSON file and images
   - JSON : filename original and new, keypoints, nb_objects, diff, size, size_class
   - Image : jpg with black pixels in the background



Approach

Object Detection Model

Pretrained MaskRCNN to detect the number of Giraffes



There are +2204 images created after pre-processing the 17 150 images with object detection model

Approach 1

Preprocessing steps

	Hypothesis	Justifications	Statistics
Image processing	H1	Remove images sizes of group [D, E]	90 images
	H2	Remove images with keypoints that doesn't belong to the giraffe object bounding box	380 images
	H3	Remove noisy and horizontal samples with ratio = too_x - toh_x / toh_y - fbh_y Keep only samples where -0.01 < ratio < 0.03	Around 3000 samples with correct ratio
	H4	Remove images where the Giraffe are cut with ratio abs(head-neck) / abs(neck-front_bottom_hoof) Keep only samples where with 0 < ratio < 0.85	-
	H5	Detect horizontal giraffes and make a rotation to have only vertical giraffes	Around 20% of horizontal images

Approach

Preprocessing steps

	Hypothesis	Justifications
Keypoints processing	Ha	• Remove images with at least one missing keypoints (total number of annotations = 0)
	Hb(x)	• Duplicate images by a factor of x and use random combination of the original annotations for each keypoint
	Hc	• Remove wrong annotations by computing the angle between toh-ni-fbh • Duplicate images by a factor of x and use random combination of the original annotations for each keypoint
	Hd	• Remove wrong annotations by taking the median of annotations per keypoints and generate keypoints accordingly to the uniform distribution in an intervall of [median-2 ; median + 2] • Duplicate images by a factor of x and use random combination of the generated annotations for each keypoint

Approach 1

Use the ResNet50 backbone

- Train : 70%, Valid : 15% and Test : 15%

Use a pretrained ResNet50 model to extract features :

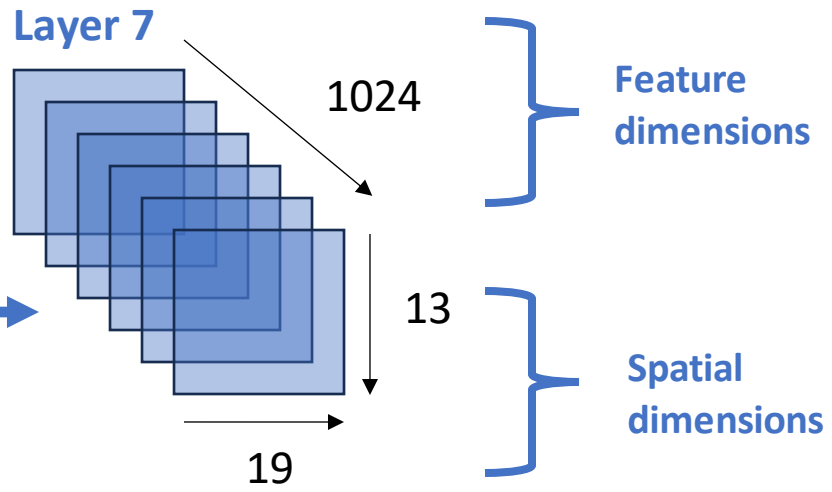
ResNet50 can extract different sizes of features :

L6: Size([N, 512, 25, 38])

L7: Size([N, 1024, 13, 19])

L8: Size([N, 2048, 7, 10])

L9: Size([N, 2048, 1, 1])



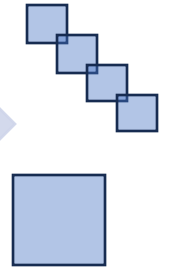
Use additional layers to process features and make predictions

Use layers to decrease features dimension

Use pooling layers to decrease spatial dimensions

Size([N, 1024, **13, 19**]) to Size([N, **1024, 1, 1**])

Size([N, **1024**, 13, 19]) to Size([N, **1, 13, 19**])



Loss Function

- Mean Square Error

Approach 2

Model configuration

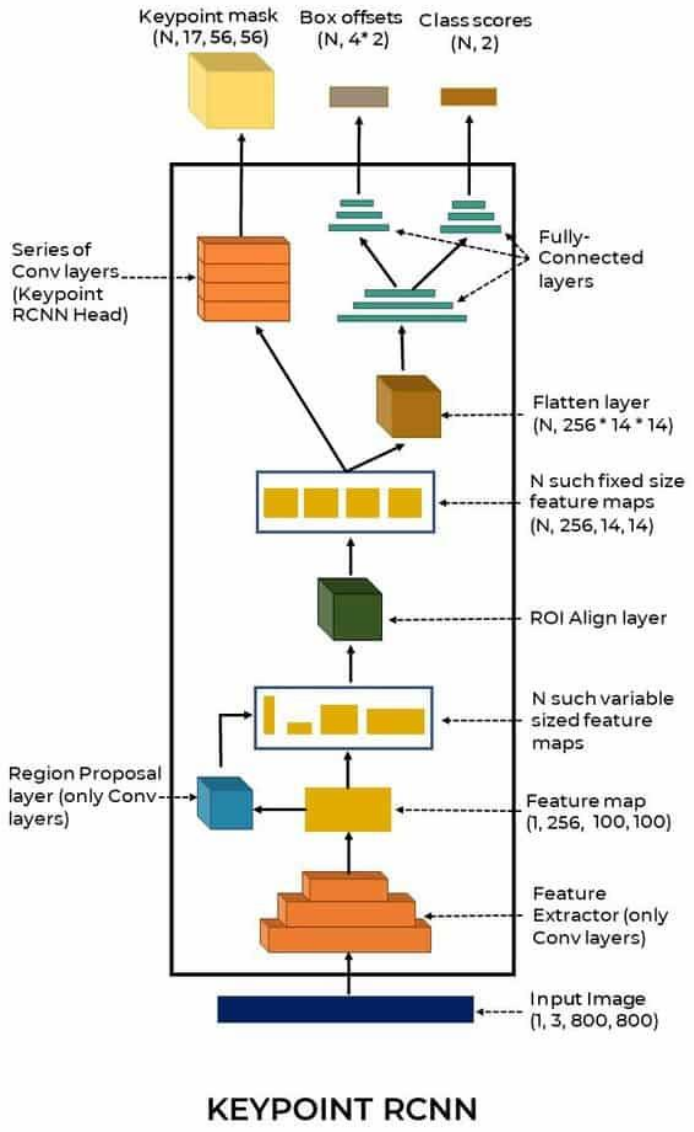
Fine tune the Keypoint-RCNN

- Train : 70%, Valid : 15% and Test : 15%

Loss Function in Keypoint-RCNN

- Input: Receive a [K, H, W] feature map.
- Ground-Truth Encoding: One-hot-encode ground-truth keypoints across K channels.
- Softmax Activation: Apply channel-wise softmax to the feature map.
- Probability Comparison: Compare softmax probabilities with ground-truth binary masks.
- Cross-Entropy Loss: Measure dissimilarity using cross-entropy loss.

$$\frac{-\sum_{h,w}\left[Y_{k,h,w}==1\right]\left(Y_{k,h,w}*\log\left(\operatorname{softmax}\left(\hat{Y}_{k,h,w}\right)\right)\right)}{\sum_{h,w}\left[Y_{k,h,w}==1\right]}$$



Results

Keypoints Detection Model

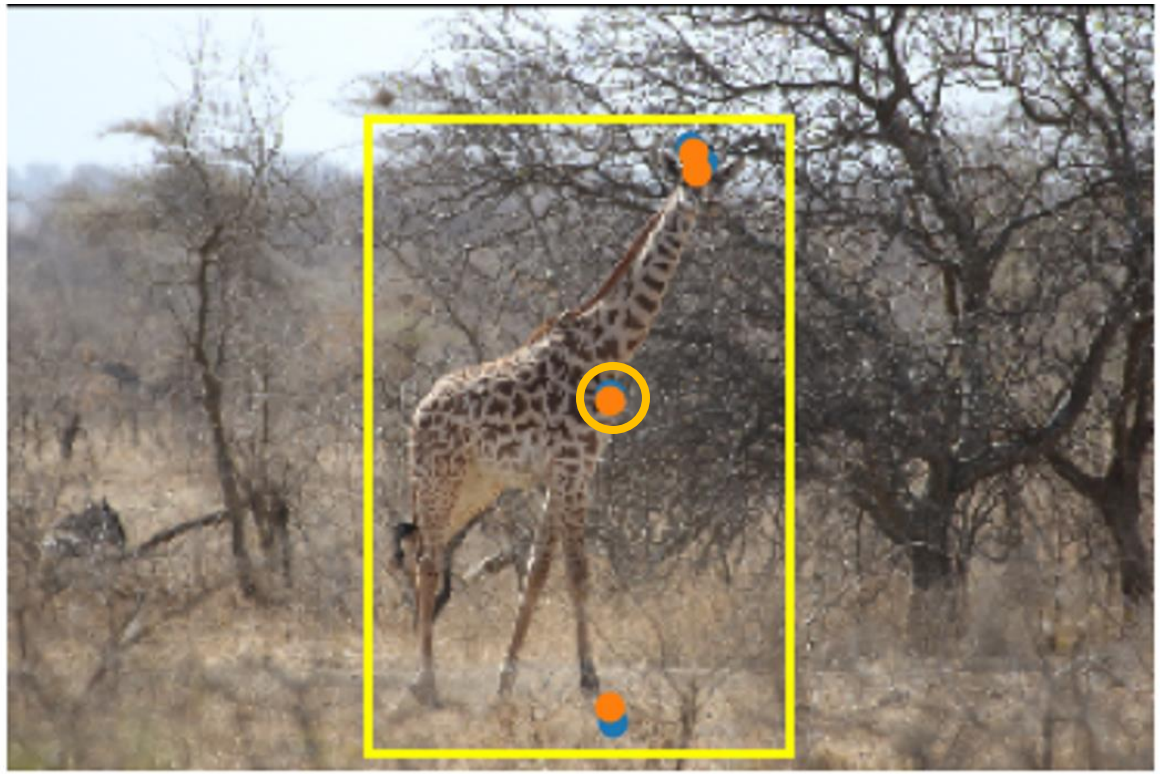
Dataset Baseline : no missing values, no images of sizes with class D or E, only one giraffe per image and remove samples where the keypoints doesn't belong to the bounding box = **1 Giraffe and Ha + H1 + H2**

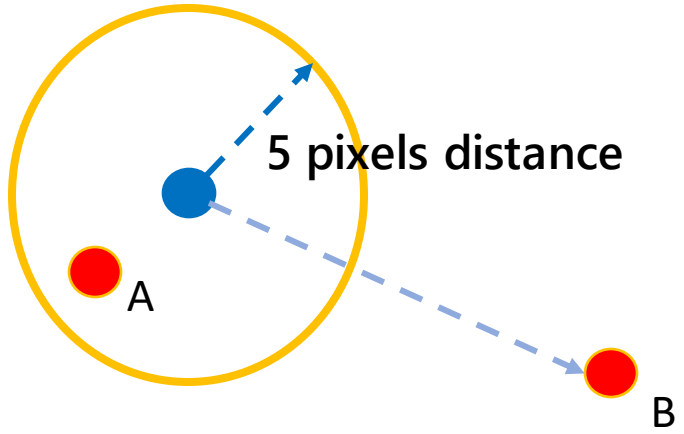
				Mean Absolute Error (pixels)			
Dataset	**Samples**	**Model**	**Aggregation**	Top of ossicone	Top of Head	Neck Indent	Front bottom hoof
Baseline + H3	-	ResNet50	Median	20.01	19.58	10.15	18.14
Baseline	10096	RCNN	Median	94.82	94.07	91.50	113.63
Baseline + H4 + H5	6147	RCNN	Median	18.83	18.94	**2.03**	7.33
Baseline + H3 + H4 + H5	2876	RCNN	Median	**3.91**	**3.9**	2.31	8.25
Baseline + Hb(2) + H3 + H4 + H5	5752	RCNN	None	5.19	4.55	3.81	**6.44**

Results

Accuracy Metric

A keypoint is well predicted if the distance between the true keypoint and the predicted one is less than 5 pixels





True keypoint

Predicted keypoint

A : correctly predicted
B : not correctly predicted

Approach

Keypoints Detection Model

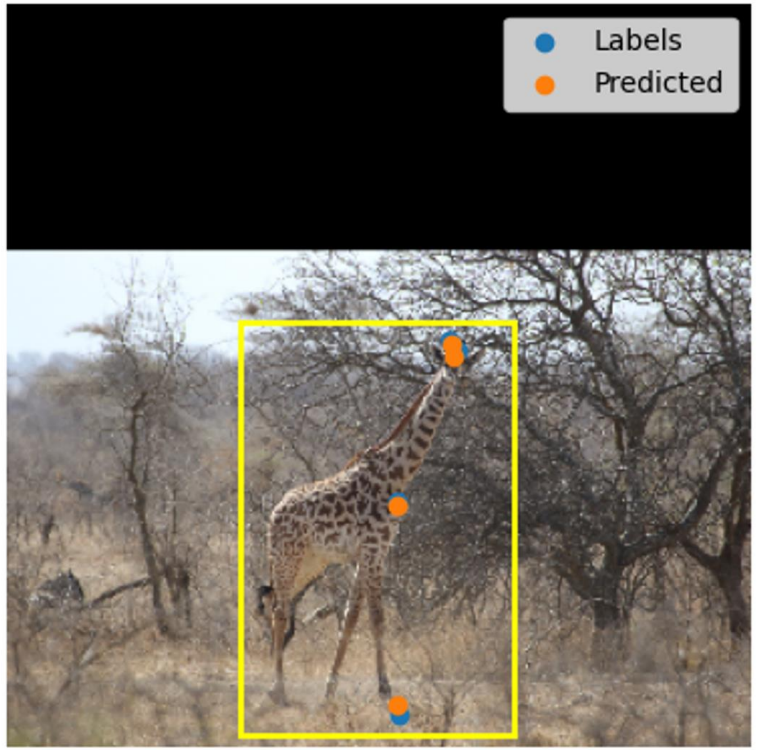
Dataset Baseline : no missing values, no images of sizes with class D or E, only one giraffe per image and remove samples where the keypoints doesn't belong to the bounding box = **1 Giraffe and Ha + H1 + H2**

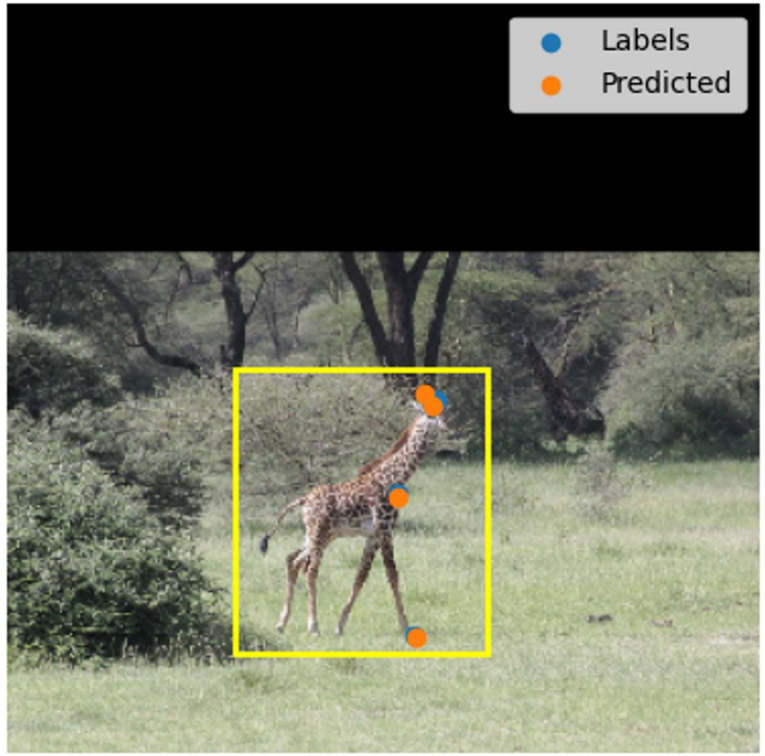
			Accuracy			
Dataset	Model	**Aggregation**	Top of ossicone	Top of Head	Neck Indent	Front bottom hoof
Baseline	RCNN	Median	4.11%	4.93%	10.96%	5.49%
Baseline + H4 + H5	RCNN	Median	63.69%	35.07%	5.30%	53.37%
Baseline + H3 + H4 + H5	RCNN	Median	80.01%	**82.73%**	**92.64%**	30.50%
Baseline + Hb(2) + H3 + H4 + H5	RCNN	None	**85.42%**	82.33%	90.26%	**65.83%**

Approach

Keypoints Detection Model

Visual results from RCNN Model

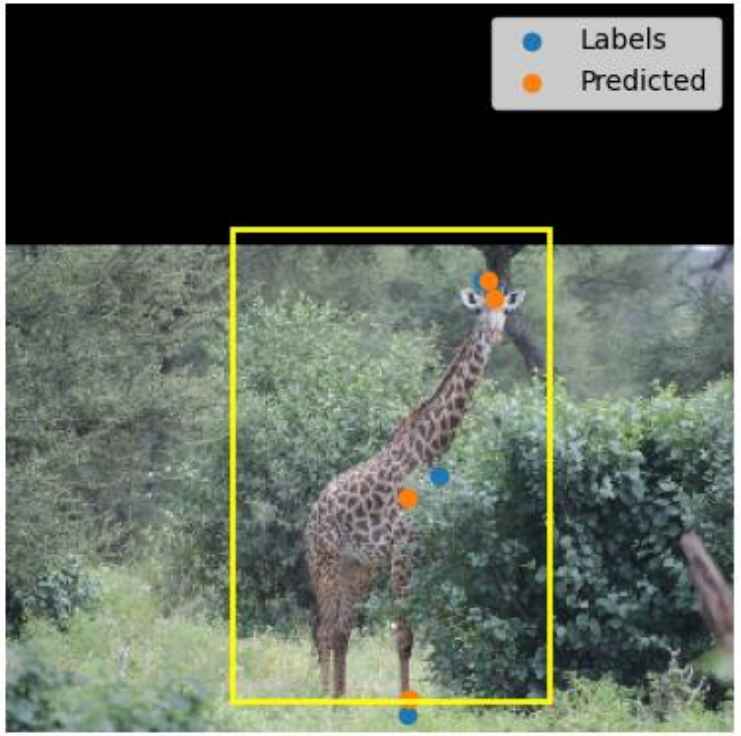


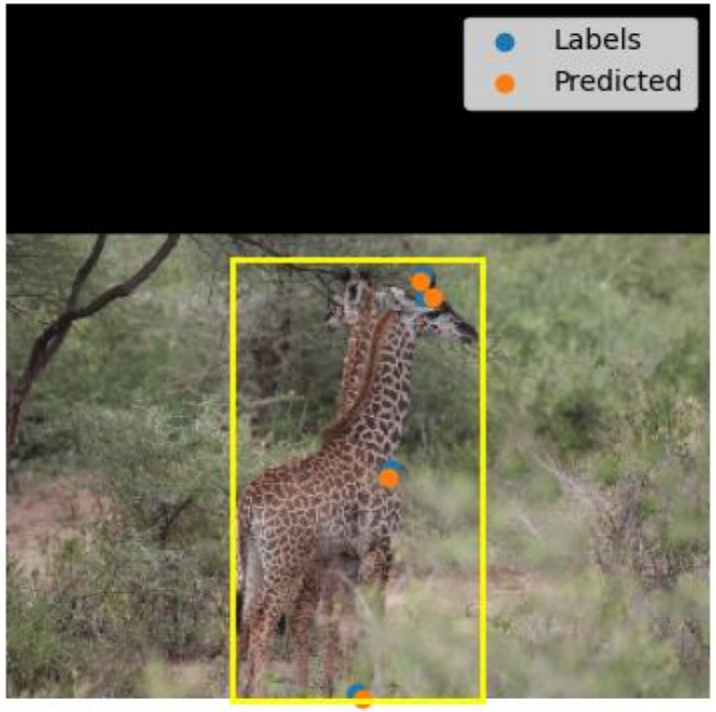


Approach

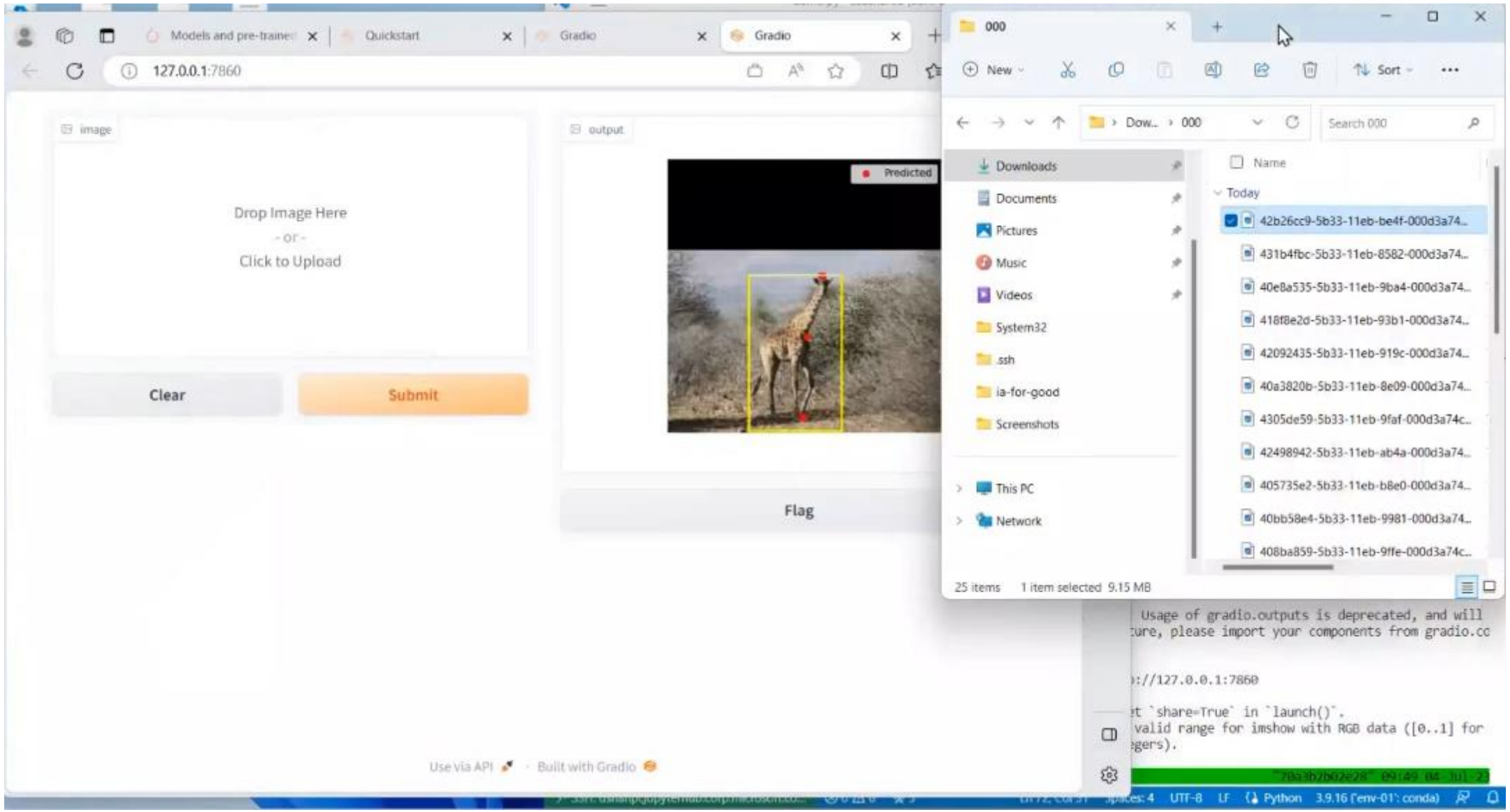
Keypoints Detection Model

Visual results from RCNN Model





Deployment Example (Gradio)



Conclusion and future work

- The quality and preprocessing of data play a crucial role in determining the success of a model
- The key point corresponding to the neck of the giraffe is easier to detect
- The ResNet50 model gives poor results even if the preprocessing steps decreases the MSE by a factor of 4.5 while the Keypoints-RCNN model gives better accuracy
- The MAE is a better metric to evaluate the model performances (MSE is really high because of outliers)
- Package the model as an executable that Derek and Monica can run on their laptops

Microsoft | AI FOR GOOD

WILD NATURE INSTITUTE

Microsoft

Thank you.

Microsoft | AI FOR GOOD

WILD NATURE INSTITUTE