

The “common thread” of the training

ANF Workflow et reproductibilité en Bioinformatique

ANF training team

The “common thread” of the training

inspired from a Nextflow training: [nf-training-intro](#)

script bash with 4 steps:

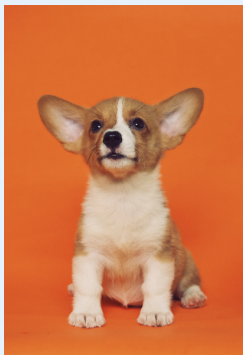
- label each image with a list of labels
- resize all images
- make a collage of the animals in each class
- combine the collages to create a single poster of the classified animals

Dataset (resized manually)

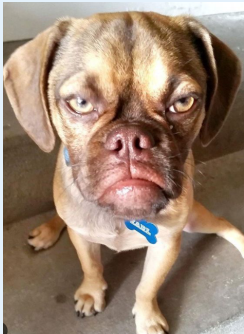
aussie



chihuahua



dog



hiding



pug



rain-ready



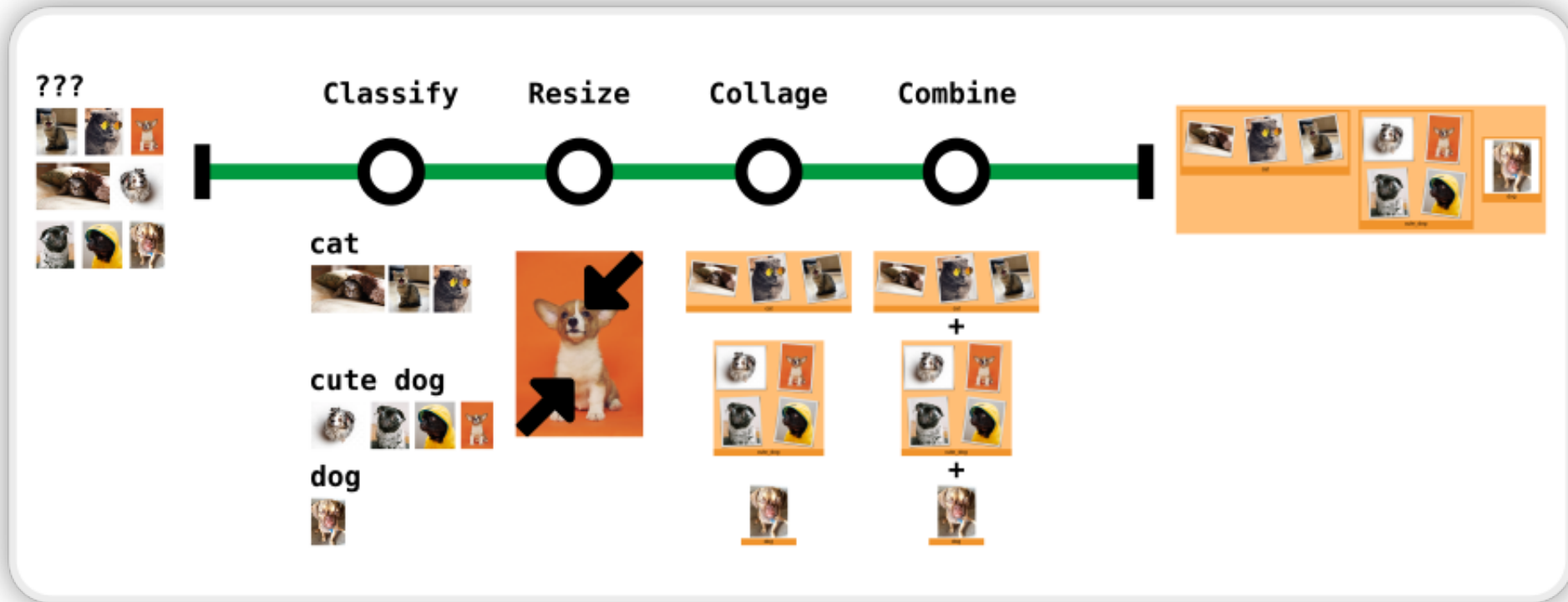
reflective



yawn



Graphical overview



third-party tools

- `classify.py`: to classe critters following the labels (use IA classification with pytorch with CLIP (Contrastive Language-Image Pre-Training) a neural network trained on a variety of (image, text) pairs (model ViT-B/32)
- `open-classify`: similar version than `classify.py` but based on FOSS (CLIP is not open)
- `mogrify` to resize images (from the ImageMagik software)
- `montage` to make the collage (from the ImageMagik software)

Initial bash script (CLIP version)

```
#!/usr/bin/env bash

# Loop over all image files in the 'data' folder to classify them
echo 'Classifying images...'
for picture in data/*; do
    # Run the classify command and fetch the label
    label=$(classify.py --image $picture --labels 'cat,dog,spider')
    # Create an output directory (folder) for this label, if it doesn't exist
    mkdir -p "classified/$label"
    # Copy the input file to the class directory
    cp $picture "classified/$label/"
done

# Resize each image
echo 'Resizing images...'
for labeldir in classified/*; do
    label=$(basename "$labeldir")
    mkdir -p "resized/$label"
```

Initial bash script (CLIP version)

```
    mogrify -resize 100x100 -path "resized/$label" -format png "$labeldir/*"
done

# Make a collage for each class
echo 'Making a collage for each label...'
mkdir -p collages
for labeldir in resized/*; do
    label=$(basename "$labeldir")
    montage -background black +polaroid -background '#ffbe76' "$labeldir/*" png:- \
    | montage -label "$label" -geometry +0+0 -background "#f0932b" - "collages/
$label.png"
done

# Combine the collages
echo 'Combining for final image...'
montage -geometry +10+10 -quality 05 -background "#ffbe76" -border 5 -bordercolor
"#f0932b" collages/* collage_all.png
echo 'Done!'
```

Runtime environment

```
ssh <my_ifb_login>@core.cluster...
cd /shared/projects/2557_anf_workflow/participants
mkdir -p <my_ifb_login> ; cd <my_ifb_login>
git clone https://github.com/seqeralabs/nf-training-intro.git
# follow the readme:
# https://github.com/seqeralabs/nf-training-intro/blob/master/docs/part2-bash.md
cd nf-training-intro/exercice/bash
module load python/3.12 imagemagick/7.1.0_5
./make_collage.sh
# => classify.py: command not found
# add "python" or in sh before call classify.py
ln -s ../../bin/classify.py
# => ModuleNotFoundError: No module named 'clip'
# no simple solution
```