

# images\_gets

August 4, 2022

## 1 Récupération d'images avec Bing

Material for the hackathon ENSAE / BRGM / 2018. Les images sont extraites de tweets mais sont retweetées sans être retweetées.

```
[1]: %matplotlib inline
import matplotlib.pyplot as plt
```

```
[2]: from jyquickhelper import add_notebook_menu
add_notebook_menu()
```

### 1.1 Récupération d'images

#### 1.1.1 Télécharger des images depuis ImageNet

Sources possibles: [ImageNet](#)

```
[3]: from ensae_projects.hackathon.image_helper import stream_download_images
dest_folder = "c:/temp/suricatenat_images/imagenet7"
res = list(stream_download_images("c:/temp/suricatenat_images/imagenet.synset7.txt",
                                dest_folder, fLOG=print, skip=100))
len(res)
```

#### 1.1.2 Télécharger des images depuis Bing

On peut s'inscrire sur l'API [Bing](#) ou télécharger quelques images. Les exemples suivant ont été récupérant en sauvant la page [inondations 2016](#) après avoir fait défiler les résultats pour en afficher beaucoup.

```
[4]: from ensae_projects.hackathon.web_search_helper import extract_bing_result
urls = extract_bing_result("c:/temp/suricatenat_clean/urls/small bridge france - Bing_
    ↪images.html")
print(len(urls))
urls[:5]
```

735

```
[4]: ['https://d2v9y0dukr6mq2.cloudfront.net/video/thumbnaill/NyjYWBnugijqylcxa/videob
locks-eiffel-tower-sunrise-timelapse-with-boats-on-seine-river-and-in-paris-
france-view-from-grenelle-bridge_sohfmyy2w_thumbnail-small01.jpg',
'https://frank.itlab.us/france_2005/small_france/dsc_1739.jpg',
'http://www.nickbooth.id.au/Europe11/images/Honfleur/H-reflects.jpg',
'http://www.tauck.com/~media/Images/Brand+Landing/Bridges/Bridges+Home+page+Ba
```

```
nner/BridgesHome_ZS.ashx',  
'https://frank.itlab.us/france_2005/small/jul_23_paris_eurosat.jpg']
```

### 1.1.3 Télécharger des images depuis une liste d'urls

```
[5]: from ensae_projects.hackathon.image_helper import stream_download_images  
got = list(stream_download_images(urls, dest_folder="c:/temp/suricatenat_clean/bing/",  
↳ fLOG=print))  
got[:5]
```

```
[stream_download_images] ... 1/735: 'https://d2v9y0dukr6mq2.cloudfront.net/video  
/thumbnail/NyjYWBnugijqylcxa/videoblocks-eiffel-tower-sunrise-timelapse-with-  
boats-on-seine-river-and-in-paris-france-view-from-grenelle-  
bridge-sohfmyy2w_thumbnail-small101.jpg'  
[stream_download_images] error 403 for url  
'http://www.nickbooth.id.au/Europe11/images/Honfleur/H-reflects.jpg'.  
[stream_download_images] error 404 for url 'http://www.tauck.com/~media/Images/  
Brand+Landing/Bridges/Bridges+Home+page+Banner/BridgesHome_ZS.ashx'.  
[stream_download_images] wrong filename for url 'https://a0.muscache.com/im/pict  
ures/0d274072-bb89-45d2-9188-20b9535a36c4.jpg?aki_policy=x_large'.  
[stream_download_images] wrong filename for url 'https://laughingsquid.com/wp-  
content/uploads/2017/09/a-logging-truck-turns-and-impressively-crosses-a-small-  
bridge-in-france-with-a-full-load.png?w=750'.  
[stream_download_images] error 503 for url 'http://stuffpoint.com/bridges-  
architectural-wonders-around-the-world/image/406101-bridges-architectural-  
wonders-around-the-world-lovely-bridge-reaching-a-small-landmass-in-the-bay-  
of.jpg'.  
[stream_download_images] wrong filename for url 'http://i41.photobucket.com/albu  
ms/e257/riruriru/Fronkreich%202013/DSC01594_small_zps6d259b99.jpg-original'.  
[stream_download_images] wrong filename for url 'https://a0.muscache.com/im/pict  
ures/76fb1fb6-1b26-4c85-9e8e-af4f13bc15e1.jpg?aki_policy=x_large'.  
[stream_download_images] error 400 for url 'https://www.wallpaperup.com/uploads/  
wallpapers/2014/03/26/309890/0e773b5b7d8aa2f0a2bf1fb339ce61ee.jpg'.  
[stream_download_images] wrong filename for url  
'http://i1.wp.com/romancolosseum.org/wp-content/uploads/2012/01/pont-du-  
gard.jpg?resize=500%2C375'.  
[stream_download_images] error 404 for url 'http://iphonewalls.net/wp-  
content/uploads/2016/03/Venice+Italy+Small+Bridge+iPhone+6+Wallpaper.jpg'.  
[stream_download_images] error 403 for url  
'http://www.planetminecraft.com/files/resource_media/screenshot/1123/canyon-  
spawn_86247.jpg'.  
[stream_download_images] ... 101/735: 'https://i.dailymail.co.uk/i/pix/2015/10/1  
3/18/2D5FF8F500000578-3270916-Originally_the_Carrick_a_Rede_Rope_Bridge_in_North  
ern_Ireland_on-m-83_1444757867849.jpg'  
[stream_download_images] cannot load image for url 'http://sca.salemsattic.com/1  
wblog/gallery/2/2007-02-09+Swiss+watching+construction+on+the+foot+bridge.jpg'  
due to cannot identify image file <_io.BytesIO object at 0x000001F8129F0FC0>  
[stream_download_images] fails for url 'https://pull-liveandinvestoverseas-  
liveandinvestove.netdna-ssl.com/wp-content/uploads/2017/03/paris-bridge-  
small.jpg' due to HTTPSConnectionPool(host='pull-liveandinvestoverseas-  
liveandinvestove.netdna-ssl.com', port=443): Max retries exceeded with url: /wp-  
content/uploads/2017/03/paris-bridge-small.jpg (Caused by  
NewConnectionError('<urllib3.connection.VerifiedHTTPSConnection object at
```

```

0x000001F812A0B780>: Failed to establish a new connection: [Errno 11001]
getaddrinfo failed')).
[stream_download_images] wrong filename for url 'https://a0.muscache.com/im/pictures/51376541/33c949fc_original.jpg?aki_policy=x_large'.
[stream_download_images] wrong filename for url
'http://i2.wp.com/traveluto.com/wp-content/uploads/2016/03/pont_du_gard.jpg?resize=810%2C536'.
[stream_download_images] wrong filename for url
'http://media.gettyimages.com/photos/construction-of-a-boat-bridge-over-the-small-rhine-french-army-in-picture-id859849280?s=612x612'.
[stream_download_images] wrong filename for url 'http://i41.photobucket.com/albums/e257/riruriru/Fronkreich%202013/DSC01604_small_zps0380043e.jpg-original'.
[stream_download_images] wrong filename for url
'https://i2.wp.com/www.wayfaringwithwagner.com/wp-content/uploads/2017/07/colmar-france-wayfaring-with-wagner-8.jpg?resize=1000%2C667&ssl=1'.
[stream_download_images] wrong filename for url
'https://render.fineartamerica.com/images/rendered/small/totebag/images/artworkimages/medium/1/bridge-of-alexandre-iii-in-paris-france-anastasy-yarmolovich.jpg?transparent=0&targetx=-181&targety=0&imagewidth=1126&imageheight=763&modelwidth=763&modelheight=763&backgroundcolor=9DA2AC&orientation=0&producttype=totebag-18-18&imageid=3275803'.
[stream_download_images] ... 201/735:
'https://www.telegraph.co.uk/content/dam/Travel/2017/February/procidaitaly.jpg?imwidth=450'
[stream_download_images] wrong filename for url
'https://www.telegraph.co.uk/content/dam/Travel/2017/February/procidaitaly.jpg?imwidth=450'.
[stream_download_images] wrong filename for url 'http://images.adsttc.com/media/images/5327/bbb2/c07a/805c/d800/033c/large_jpg/IMG_9818-9.jpg?1395112873'.
[stream_download_images] error 404 for url
'http://www.ventanasvoyage.com/images/France+selects/bridge_small.jpg'.
[stream_download_images] wrong filename for url 'http://cdn.newsapi.com.au/image/v1/1cc6dfda05ccc935fdac0b9157038216?width=1024'.
[stream_download_images] error 404 for url
'http://www.preservationpa.org/uploads/historic+properties+for+sale/Compass-Mill_Lititz_for-sale_R.Harris.jpg'.
[stream_download_images] wrong filename for url 'https://images.adsttc.com/media/images/58ed/1d1e/e58e/ce8f/0300/030e/large_jpg/FRONT_RENDERING_2.jpg?1491934486'.
[stream_download_images] error 404 for url 'https://www.orbitz.com/blog/wp-content/uploads/2015/08/Annecy.jpg'.
[stream_download_images] wrong filename for url
'https://i1.wp.com/travelsfrance.com/wp-content/uploads/2016/05/Carbonniere-Tower-Camargue-B.jpg?fit=1064%2C710&ssl=1'.
[stream_download_images] wrong filename for url 'https://media-cdn.sygictraveldata.com/media/800x600/612664395a40232133447d33247d3835333831393034'.
[stream_download_images] cannot load image for url 'http://sca.salemsattic.com/1wblog/gallery/2/2007-02-09+foot+bridge+over+the+Rhein+from+Germany+to+France+shore.jpg' due to cannot identify image file <_io.BytesIO object at 0x000001F812D22CA8>
[stream_download_images] wrong filename for url 'https://s-media-cache-ak0.pinning.com/236x/79/b8/e6/79b8e64e1854dd69a04bb276fc0bc73f.jpg?noindex=1'.
[stream_download_images] error 400 for url 'https://www.wallpaperup.com/uploads/

```

wallpapers/2013/02/15/40252/c0906da8d3d30169c38c44c7cb07de2c.jpg' .  
[stream\_download\_images] wrong filename for url 'http://i41.photobucket.com/albums/e257/riruriru/Fronkreich%202013/DSC01596\_small\_zpsa8d0f4b3.jpg-original' .  
[stream\_download\_images] ... 301/735:  
'https://render.fineartamerica.com/images/rendered/small/poster/images-square-real-5/1-valentre-bridge-in-cahors-france-elena-elisseeva.jpg'  
[stream\_download\_images] error 400 for url 'https://www.wallpaperup.com/uploads/wallpapers/2014/12/21/560794/1ac7056f6b90748b083fba5cbee8f601-700.jpg' .  
[stream\_download\_images] wrong filename for url 'https://cdn.theatlantic.com/assets/media/img/2015/07/29/RTR4ILKC/1920.jpg?1438183123' .  
[stream\_download\_images] wrong filename for url  
'https://qph.fs.quoracdn.net/main-qimg-164aa27c9198a9eff4c036c56849340e-c' .  
[stream\_download\_images] wrong filename for url 'https://sheerluxe.com/sites/default/files/styles/sl\_free\_responsive/public/media/2018/11/provision.jpg?itok=Zf0\_9gsd' .  
[stream\_download\_images] wrong filename for url 'https://www.vangoghmuseum.nl/download/c82c2fcd-90e8-4de6-b23d-f28c2af907d6.jpg?size=m' .  
[stream\_download\_images] wrong filename for url  
'http://media.gettyimages.com/photos/small-boat-passes-under-pegasus-bridge-normandy-france-europe-site-of-picture-id599378690?s=612x612' .  
[stream\_download\_images] wrong filename for url 'https://surfaceview.s3.amazonaws.com/spree/products/187/small/TRU0016-AP-E.jpg?1408016791' .  
[stream\_download\_images] wrong filename for url  
'http://media.gettyimages.com/photos/normandy-france-picture-id166836920?s=170667a' .  
[stream\_download\_images] wrong filename for url 'https://a0.muscache.com/im/pictures/e5a8895a-47c0-40eb-a8ae-94cfd5039607.jpg?aki\_policy=x\_large' .  
[stream\_download\_images] ... 401/735:  
'http://www.eutouring.com/place\_dauphine\_m13\_DSC01821FP\_lrg.JPG'  
[stream\_download\_images] wrong filename for url  
'https://qph.ec.quoracdn.net/main-qimg-ff7fe9b391213392474d394c3245f3ca-c?convert\_to\_webp=true' .  
[stream\_download\_images] wrong filename for url 'https://surfaceview.s3.amazonaws.com/spree/products/179/small/BLI0124-AP-E.jpg?1408016006' .  
[stream\_download\_images] fails for url 'https://frenchmoments.eu/wp-content/uploads/2015/05/Pont-de-la-Concorde-04-©-French-Moments.jpg' due to 'ascii' codec can't encode character '\xa9' in position 55: ordinal not in range(128) .  
[stream\_download\_images] cannot load image for url 'http://www.mountain-lifestyle.com/images/graphics/small/ski-area-planards-pistes-winter-chamonix-01.jpg' due to cannot identify image file <\_io.BytesIO object at 0x000001F812D22990>  
[stream\_download\_images] wrong filename for url 'http://il4.picdn.net/shutterstock/videos/4559468/thumb/1.jpg?i10c=img.resize(height:160)' .  
[stream\_download\_images] wrong filename for url  
'https://hikeminded.files.wordpress.com/2018/11/dsc\_0240.jpg?w=1200' .  
[stream\_download\_images] error 404 for url 'http://www.gagnac-sur-cere.com/photos/Millau+Viaduct.jpg' .  
[stream\_download\_images] ... 501/735:  
'https://i.pinimg.com/736x/95/7f/d3/957fd3418f8bf8bfc09dd17eaac063c5--garden-ponds-koi-ponds.jpg'  
[stream\_download\_images] wrong filename for url  
'https://render.fineartamerica.com/images/rendered/small/flat/weekender-tote-bag/images-medium-5/2-valentre-bridge-in-cahors-france-elena-elisseeva.jpg?trans

parent=0&targetx=0&targety=-328&imagewidth=779&imageheight=1163&modelwidth=779&modelheight=506&backgroundcolor=28190E&orientation=0&producttype=totebagweekender-24-16-white'.

[stream\_download\_images] wrong filename for url 'https://render.fineartamerica.com/images/rendered/small/shower-curtain/images-medium-5/1-valentre-bridge-in-cahors-france-elena-elisseeva.jpg?transparent=0&targetx=-218&targety=0&imagewidth=1223&imageheight=819&modelwidth=787&modelheight=819&backgroundcolor=CADFF3&orientation=0&producttype=showercurtain'.

[stream\_download\_images] wrong filename for url 'http://media.gettyimages.com/photos/small-bridge-over-a-canal-by-the-chateau-de-chenonceau-on-the-river-picture-id558031533?s=594x594'.

[stream\_download\_images] wrong filename for url 'https://render.fineartamerica.com/images/rendered/small/greeting-card/images/artworkimages/medium/1/old-stone-bridge-in-france-menega-sabidussi.jpg?transparent=0&targetx=0&targety=-12&imagewidth=700&imageheight=525&modelwidth=700&modelheight=500&backgroundcolor=70500B&orientation=0&producttype=greetingcard&imageid=50379'.

[stream\_download\_images] wrong filename for url 'https://smallhousebliss.files.wordpress.com/2015/03/covert-cabin-woodsman-exterior4-via-smallhousebliss.jpg?w=1200'.

[stream\_download\_images] ... 601/735: 'http://n450v.alamy.com/450v/bgx4e4/woman-bicycle-small-bridge-canal-du-midi-trebes-by-carcassonne-aude-bgx4e4.jpg'

[stream\_download\_images] wrong filename for url 'http://images.mentalfloss.com/sites/default/files/istock-496707858.jpg?resize=1100x740'.

[stream\_download\_images] wrong filename for url 'https://render.fineartamerica.com/images/rendered/small/flat/pouch/images-medium-5/2-valentre-bridge-in-cahors-france-elena-elisseeva.jpg?transparent=0&targetx=0&targety=-343&imagewidth=777&imageheight=1160&modelwidth=777&modelheight=474&backgroundcolor=28190E&orientation=0&producttype=pouch-regularbottom-medium'.

[stream\_download\_images] wrong filename for url 'https://render.fineartamerica.com/images/rendered/small/flat/weekender-tote-bag/images-medium-5/1-valentre-bridge-in-cahors-france-elena-elisseeva.jpg?transparent=0&targetx=0&targety=-7&imagewidth=779&imageheight=521&modelwidth=779&modelheight=506&backgroundcolor=CADFF3&orientation=0&producttype=totebagweekender-24-16-white'.

[stream\_download\_images] wrong filename for url 'https://cdn-image.travelandleisure.com/sites/default/files/styles/1600x1000/public/1465330137/Infinity-Pool-Viking-CRUISE0516.jpg?itok=2TUixPrV'.

[stream\_download\_images] fails for url 'https://frenchmoments.eu/wp-content/uploads/2015/07/Quais-de-la-Seine-©-French-Moments-Paris-34.jpg' due to 'ascii' codec can't encode character '\xa9' in position 50: ordinal not in range(128).

[stream\_download\_images] wrong filename for url 'https://assets.atlasobscura.com/media/W1siZiIsInVwbG9hZHMvcGxhY2VfaW1hZ2VzLzQ2ZjRjMzZkZjk1NWE1OTczOF80Mjg5NTUxMDE3XzQ3MjBlZTdmMTFFfYi5qcGciXSxbInAiLCJ0aHVtYiIsIngzOTA-IloSWyJwiIwiY29udmVydCIiIixdWFsaXR5IDgxIC1hdXRvLW9yaWVudCJdXQ'.

[stream\_download\_images] error 400 for url 'https://www.wallpaperup.com/uploads/wallpapers/2014/12/21/560794/1ac7056f6b90748b083fba5cbee8f601.jpg'.

[stream\_download\_images] wrong filename for url 'https://laughingsquid.com/wp-content/uploads/2017/09/a-logging-truck-turns-and-impressively-crosses-a-small-bridge-in-france-with-a-full-load.gif?w=750'.

[stream\_download\_images] ... 701/735: 'https://2.bp.blogspot.com/-7wREwz142q4/V9awba-EwFI/AAAAAAAAIAs/QhvzwWEyMfGb6\_7N

```

UtWjdNEZ3k2Csq8GwCEw/s1600/Spring%2Bon%2Bthe%2BRidge%2B1200.jpg'
[stream_download_images] wrong filename for url 'http://newimages.yachtworld.com
/resize/1/40/38/4904038_20141231070343240_1_XLARGE.jpg?f=/1/40/38/4904038_201412
31070343240_1_XLARGE.jpg&w=5315&h=2762&t=1420009880000'.
[stream_download_images] wrong filename for url
'https://frankeber.files.wordpress.com/2011/08/pont-des-arts-
iledelacitc3a9-2011-by-frankeber.jpg?w=625'.
[stream_download_images] error 403 for url
'http://www.infrancia.org/parigi/paris/paris/images/pontdelaseine_mirabeau.jpg'.
[stream_download_images] wrong filename for url
'http://www.driverguidefrance.com/sites/default/files/styles/slide-tour_960x360/
public/Chinon%20bridge%20Private%20guided%20tours%20to%20the%20Loire%20valley%20
by%20car%20or%20minibus%20from%20Paris%20for%20individuals%20or%20small%20groups
%20by%20Driver%20Guide%20France.jpg?itok=A5IveabR'.
[stream_download_images] error 400 for url 'http://www.fastdates.com/PitBoardFea
tures/2005EdelweissFrance/France06.050bridge864.jpg'.
[stream_download_images] fails for url 'http://traveltalesoflife.com/wp-
content/uploads/2015/04/img_6500-900x675.jpg' due to
HTTPConnectionPool(host='127.0.0.1', port=80): Max retries exceeded with url: /
(Caused by NewConnectionError('<urllib3.connection.HTTPConnection object at
0x000001F8129E9358>: Failed to establish a new connection: [WinError 10061]
Aucune connexion n'a pu être établie car l'ordinateur cible l'a expressément
refusée')).
[stream_download_images] error 400 for url 'https://www.wallpaperup.com/uploads/
wallpapers/2014/12/21/560794/1ac7056f6b90748b083fba5cbee8f601-500.jpg'.
[stream_download_images] wrong filename for url 'https://i0.wp.com/www.annes-
de-pelerinage.com/wp-content/uploads/2017/07/colmar-france.jpg?fit=1200%2C800'.
[stream_download_images] wrong filename for url
'https://i2.wp.com/www.wayfaringwithwagner.com/wp-
content/uploads/2017/07/colmar-france-wayfaring-with-wagner-8.jpg?ssl=1'.

```

```

[5]: ['videoblocks-eiffel-tower-sunrise-timelapse-with-boats-on-seine-river-and-in-
paris-france-view-from-grenelle-bridge_sohfmyy2w_thumbnail-small01.jpg',
'dsc_1739.jpg',
'jul_23_paris_eurosat.jpg',
'Bridge_Stowe_Landscape_Gardens_BY_ROBERT_KILPIN.jpg',
'falais_03.jpg']

```

#### 1.1.4 Echantillon aléatoire

```

[6]: from ensae_projects.hackathon.image_helper import last_element, stream_random_sample
rnd = last_element(stream_random_sample("c:/temp/suricatenat_images/"))
rnd[:5]

```

```

[7]: len(rnd)

```

```

[8]: from ensae_projects.hackathon.image_helper import plot_gallery_random_images
ax, rnd = plot_gallery_random_images("c:/temp/suricatenat_images/")
ax;

```

```

[9]: rnd

```

```
[10]: ax, rnd = plot_gallery_random_images("c:/temp/suricatenat_images/imagenet4/")
ax;
```

```
[11]: rnd
```

### 1.1.5 Renommer les images de l'échantillon

```
[12]: from ensae_projects.hackathon.image_helper import enumerate_image_class
ech = list(enumerate_image_class("c:/temp/sample_labelled/"))
```

```
[13]: len(ech)
```

```
[13]: 309
```

```
[14]: import os
echnames = set(os.path.split(n[0])[-1] for n in ech)
list(sorted(echnames))[:5]
```

```
[14]: ['1051720569702555648_Dph2SGiXUAAImp.jpg',
'1051787907109974016_Dpizg3kW4AAAdpaM.jpg',
'1051812613062098946_DpjJ5NWXUAAAtfer.jpg',
'1051914579549282309_DpkmtpOX4AAunhI.jpg',
'106994_5349_big_200907_voyager11.jpg']
```

```
[15]: images = list(enumerate_image_class("c:/temp/suricatenat_clean/"))
```

```
[16]: rename = []
for img, sub in images:
    name = os.path.split(img)[-1]
    if name in echnames:
        rename.append(img)
rename[:5]
```

```
[16]: ['c:/temp/suricatenat_clean/bing\\1200px-Chateau_de_Chenonceau_2008E.jpg',
'c:/temp/suricatenat_clean/bing\\1970699-presse-papiers-1.jpg',
'c:/temp/suricatenat_clean/bing\\2243060331_6496d792a2_o_d.jpg',
'c:/temp/suricatenat_clean/bing\\28080031282_6d00c773d6_b.jpg',
'c:/temp/suricatenat_clean/bing\\28183949335_04578a6bdd_b.jpg']
```

```
[17]: len(rename)
```

```
[17]: 306
```

```
[18]: for img in rename:
    os.rename(img + "~", img + ".ech")
```

### 1.1.6 Split train test

```
[19]: from ensae_projects.hackathon.image_helper import folder_split_train_test
tr, te = folder_split_train_test("c:/temp/sample_labelled/",
                                "c:/temp/sample_labelled_train/",
                                "c:/temp/sample_labelled_test/",
```

```
len(tr), len(te)                                test_size=0.66)
```

[19]: (105, 204)

[20]: