

# 指定テキスト p.74~75 2.10.3~2.10.5 でのデータ準備

<https://nihcc.app.box.com/v/ChestXray-NIHCC/>

NIH Clinical Center  
America's Research Hospital

ログイン    サインアップ

CXR8    ダウンロード

名前	サイズ	詳細
images	14ファイル	フォルダのプロパティ
README_CHESTXRAY.pdf	828.4 KB	所有者 Summers, Ronald (NIH/CC/DRD) [E]
FAQ_CHESTXRAY.pdf	2017年12月15日、更新者: 以... 70.5 KB	Enterpriseの所有者 National Institutes of Health - Clinical Center
train_val_list.txt	2017年12月15日、更新者: 以... 1.4 MB	作成日 2017年9月2日 3:33
test_list.txt	2017年12月15日、更新者: 以... 424.9 KB	変更日: 2018年12月21日 5:15
ARXIV_V5_CHESTXRAY.pdf	2017年12月15日、更新者: 以... 8.5 MB	サイズ 45.7 GB
LOG_CHESTXRAY.pdf	2017年12月15日、更新者: 以... 3.8 KB	
Data_Entry_2017.csv	2017年11月23日、更新者: 以... 7.5 MB	
BBox_List_2017.csv	2017年9月5日、更新者: 以前の... 90.3 KB	

まず、この部分をダウンロード

# imagesフォルダ内へ入る・・・

← → ↻ [nihcc.app.box.com/v/ChestXray-NIHCC](https://nihcc.app.box.com/v/ChestXray-NIHCC) ★ 👤 ⋮

**NIH Clinical Center**  
America's Research Hospital

ログイン サインアップ

CXR8 ダウンロード

次に、imagesフォルダをダブルクリックして開く。

名前	サイズ	詳細
 images	12月 21日、更新者: Ke ... 14ファイル	フォルダのプロパティ
 README_CHESTXRAY.pdf	2018年11月6日、更新者: Sum... 828.4 KB	所有者 Summers, Ronald (NIH/CC/DRD) [E]
 FAQ_CHESTXRAY.pdf	2017年12月15日、更新者: 以... 70.5 KB	Enterpriseの所有者 National Institutes of Health - Clinical Center
 train_val_list.txt	2017年12月15日、更新者: 以... 1.4 MB	作成日 2017年9月2日 3:33
 test_list.txt	2017年12月15日、更新者: 以... 424.9 KB	変更日: 2018年12月21日 5:15
 ARXIV_V5_CHESTXRAY.pdf	2017年12月15日、更新者: 以... 8.5 MB	サイズ 45.7 GB
 LOG_CHESTXRAY.pdf	2017年12月15日、更新者: 以... 3.8 KB	
 Data_Entry_2017.csv	2017年11月23日、更新者: 以... 7.5 MB	
 BBox_List_2017.csv	2017年9月5日、更新者: 以前の... 90.3 KB	

# 「名前」部分をクリックして名前順にソートする

CXR8 > images

名前

CXR8 > images

名前 ^	更新日	サイズ
 batch_download_zips.py	2018年12月21日、更新者: Ke...	1.3 KB
 images_001.tar.gz	2017年9月5日、更新者: 以前...	1.9 GB
 images_002.tar.gz	2017年9月5日、更新者: 以前...	3.7 GB
 images_003.tar.gz	2017年9月5日、更新者: 以前...	3.7 GB
 images_004.tar.gz	2017年9月8日、更新者: 以前...	3.6 GB
 images_005.tar.gz	2017年9月5日、更新者: 以前...	3.7 GB
 images_006.tar.gz	2017年9月5日、更新者: 以前...	3.7 GB
 images_007.tar.gz	2017年9月7日、更新者: 以前...	3.7 GB
 images_008.tar.gz	2017年9月5日、更新者: 以前...	3.7 GB
 images_009.tar.gz	2017年9月5日、更新者: 以前...	3.8 GB
 images_010.tar.gz	2017年9月5日、更新者: 以前...	3.9 GB
 images_011.tar.gz	2017年9月5日、更新者: 以前...	3.9 GB
 images_012.tar.gz	2017年9月5日、更新者: 以前...	2.7 GB

<https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/37178474737?sortColumn=name&sortDirection=DESC>

# 「.tar.gz」ファイルを一つずつ処理する

CXR8 > images

名前 ^



batch\_download\_zips.py



images\_001.tar.gz



images\_002.tar.gz



images\_003.tar.gz

2017年9月5日、更新者: 以前...

3.7 GB



images\_004.tar.gz

2017年9月8日、更新者: 以前...

3.6 GB



images\_005.tar.gz

2017年9月5日、更新者: 以前...

3.7 GB



images\_006.tar.gz

2017年9月5日、更新者: 以前...

3.7 GB



images\_007.tar.gz

2017年9月7日、更新者: 以前...

3.7 GB



images\_008.tar.gz

2017年9月5日、更新者: 以前...

3.7 GB



images\_009.tar.gz

2017年9月5日、更新者: 以前...

3.8 GB



images\_010.tar.gz

2017年9月5日、更新者: 以前...

3.9 GB



images\_011.tar.gz

2017年9月5日、更新者: 以前...

3.9 GB



images\_012.tar.gz

2017年9月5日、更新者: 以前...

2.7 GB

まず「Images\_001.tar.gz」  
をダウンロードする。  
1.9GB  
4,999ファイル入っている。

# 補遺：「.tar.gz」ファイルを解凍するソフトが無い場合

- 下記のどちらかから、「7-Zip」ソフトをダウンロードし、インストールしてください。
  - <https://forest.watch.impress.co.jp/library/software/7zip/>
  - <https://sevenzip.osdn.jp/>

← → ↻ [sevenzip.osdn.jp](https://sevenzip.osdn.jp)



## 7-Zip

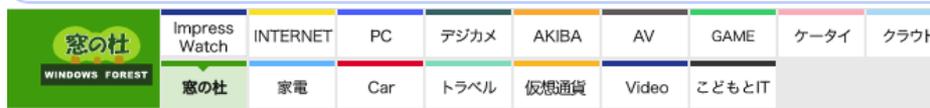
7-Zipは高圧縮率のファイルアーカイバ(圧縮・展開/圧縮・解凍ソフト)です。

7-Zip 19.00(2019-02-21) for Windowsをダウンロード:

リンク	タイプ	Windows	サイズ
<a href="#">ダウンロード</a>	.exe	32ビット x86	1 MB
<a href="#">ダウンロード</a>	.exe	64ビット x64	1 MB

ホーム  
7zファイル  
LZMA SDK  
ダウンロード  
FAQ  
サポート

forest.watch.impress.co.jp/library/software/7zip/



TOP > 圧縮・解凍・ランタイム > 圧縮・解凍 > 圧縮・解凍 > 7-Zip

## 7-Zip

**7-Zip (32bit版)**  
v19.00 (19/02/21)

7z形式などに対応した解凍・圧縮ソフト

[フリーソフト](#)

対応環境 : Windows NT/2000/XP/Server 2003/Vista/Server 2008/7/8/Server 2012/10/Server 2016

窓の社から  
ダウンロード

ファイルサイズ 1.1MB

**7-Zip (64bit版)**  
v19.00 (19/02/21)

7z形式をはじめとする書庫ファイルを圧縮・解凍するた

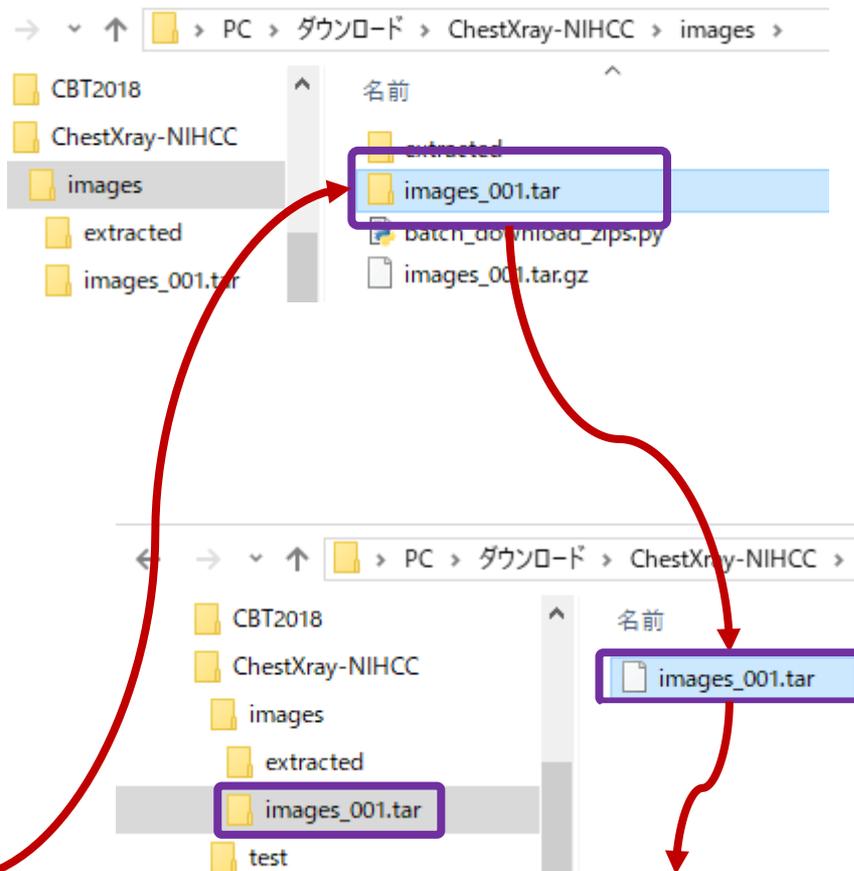
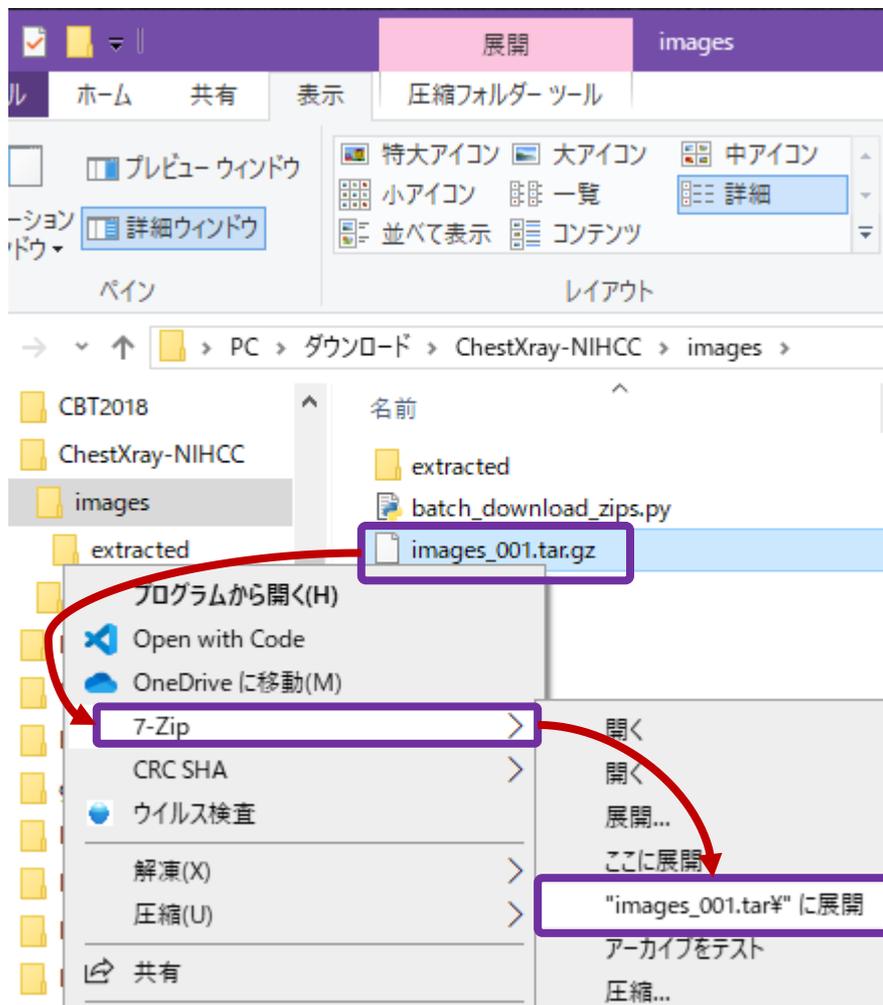
窓の社から  
ダウンロード

ファイルサイズ 1.3MB

[1]から[10]まで、###には「001」から「012」までを順に入れて繰り返す

[1] 該当する「.tar.gz」ファイルを右クリックし、[7-Zip]から[“images\_###.tar¥”に展開]を選択。

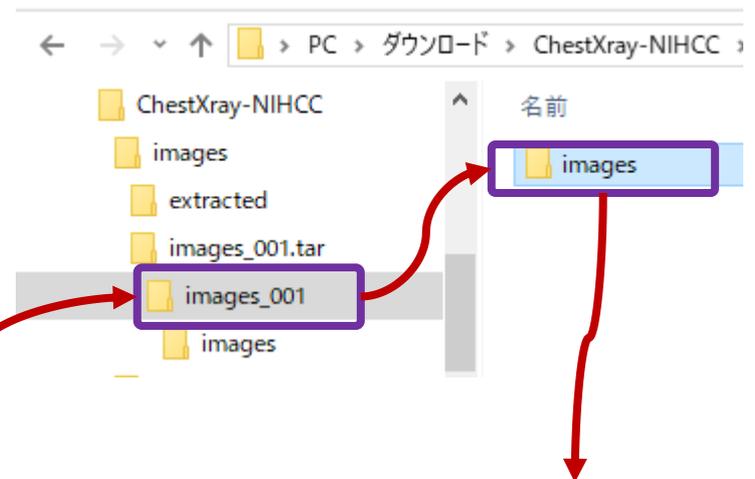
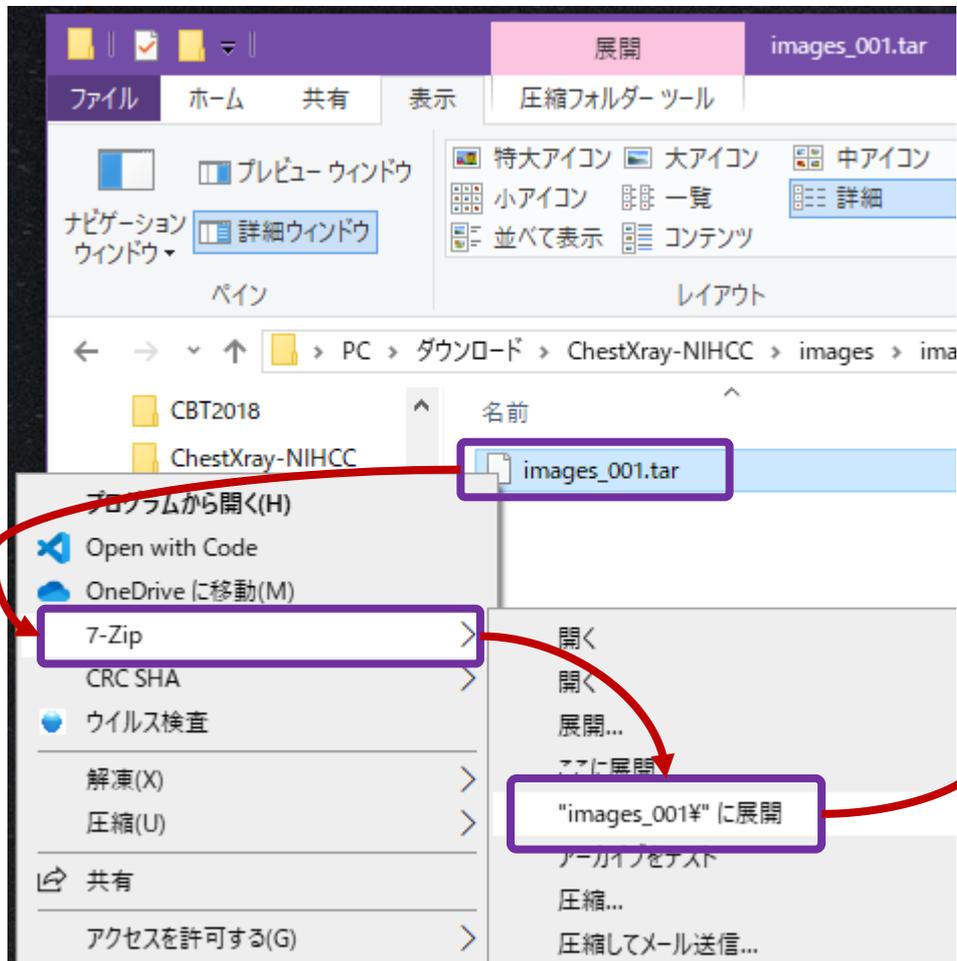
※展開後、「images\_###.tar.gz」ファイルは削除可能



この「.tar」ファイルの解凍は次ページへ...

[2] 該当する「.tar」ファイルを右クリックし、[7-Zip]から[“images\_###¥”に展開]を選択。

※展開後、「images\_###.tar」ファイルは削除可能



この「images」フォルダ内に画像がある…

# [3] 画像ファイル 1024x1024 pixel のPNGファイル

## 2段階の解凍まとめ

(各解凍後、該当する圧縮ファイルは削除可能)

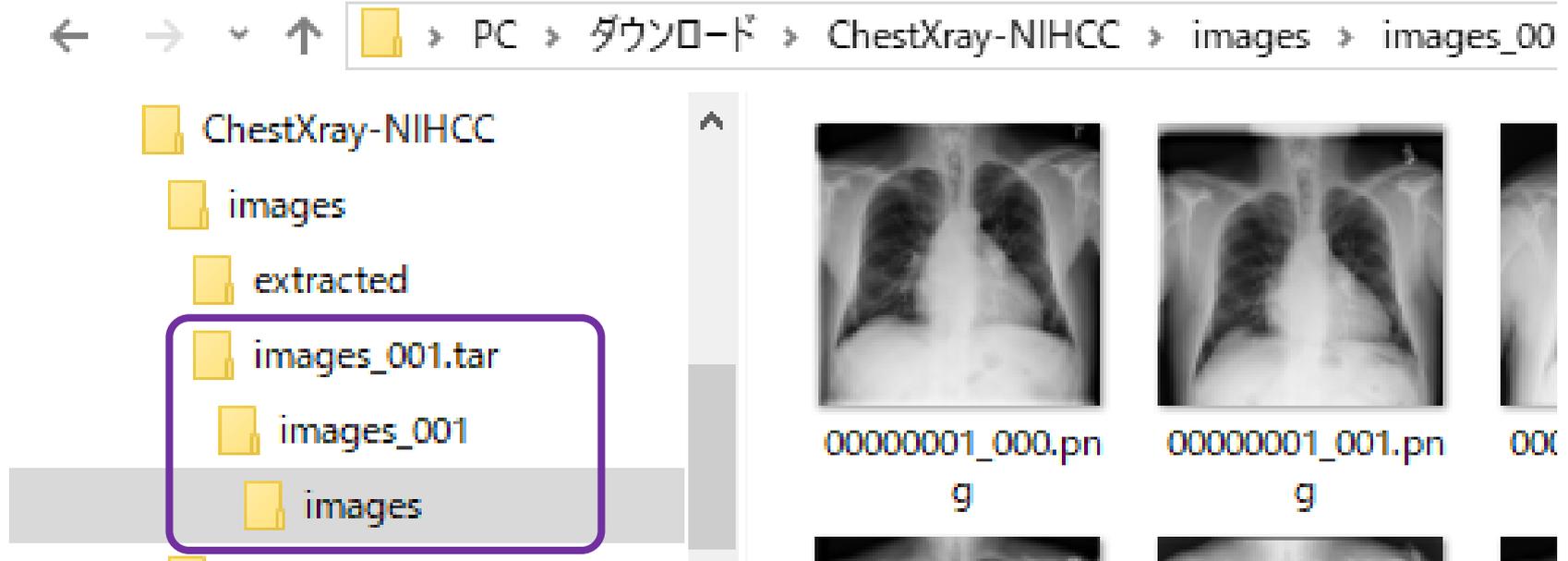
images\_###.tar.gz

→【解凍】→

[images\_###.tar] フォルダ内の images\_###.tar

→【解凍】→

[images\_###] フォルダ内の [images]フォルダに、画像ファイルがある。

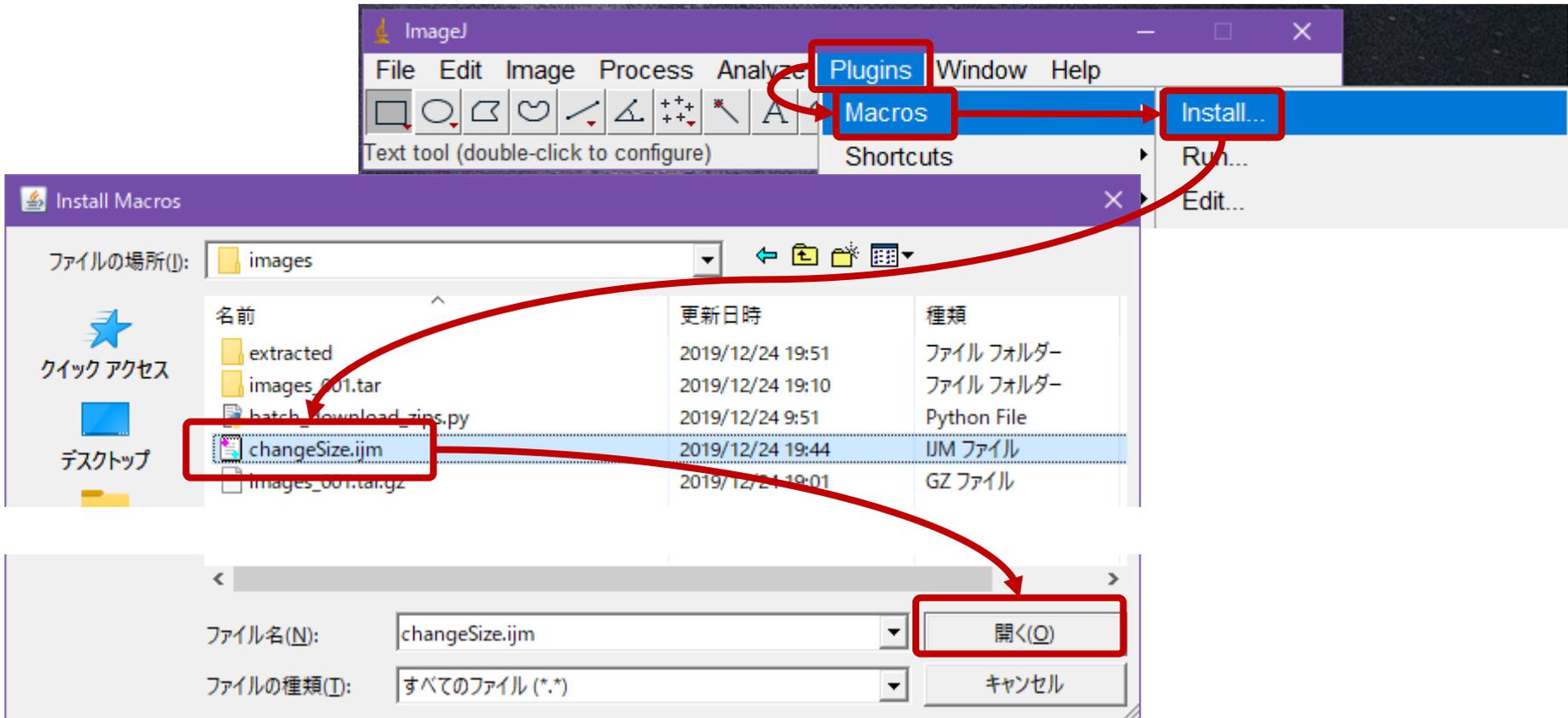


# 補遺・ImageJ ないし Fijiの インストール

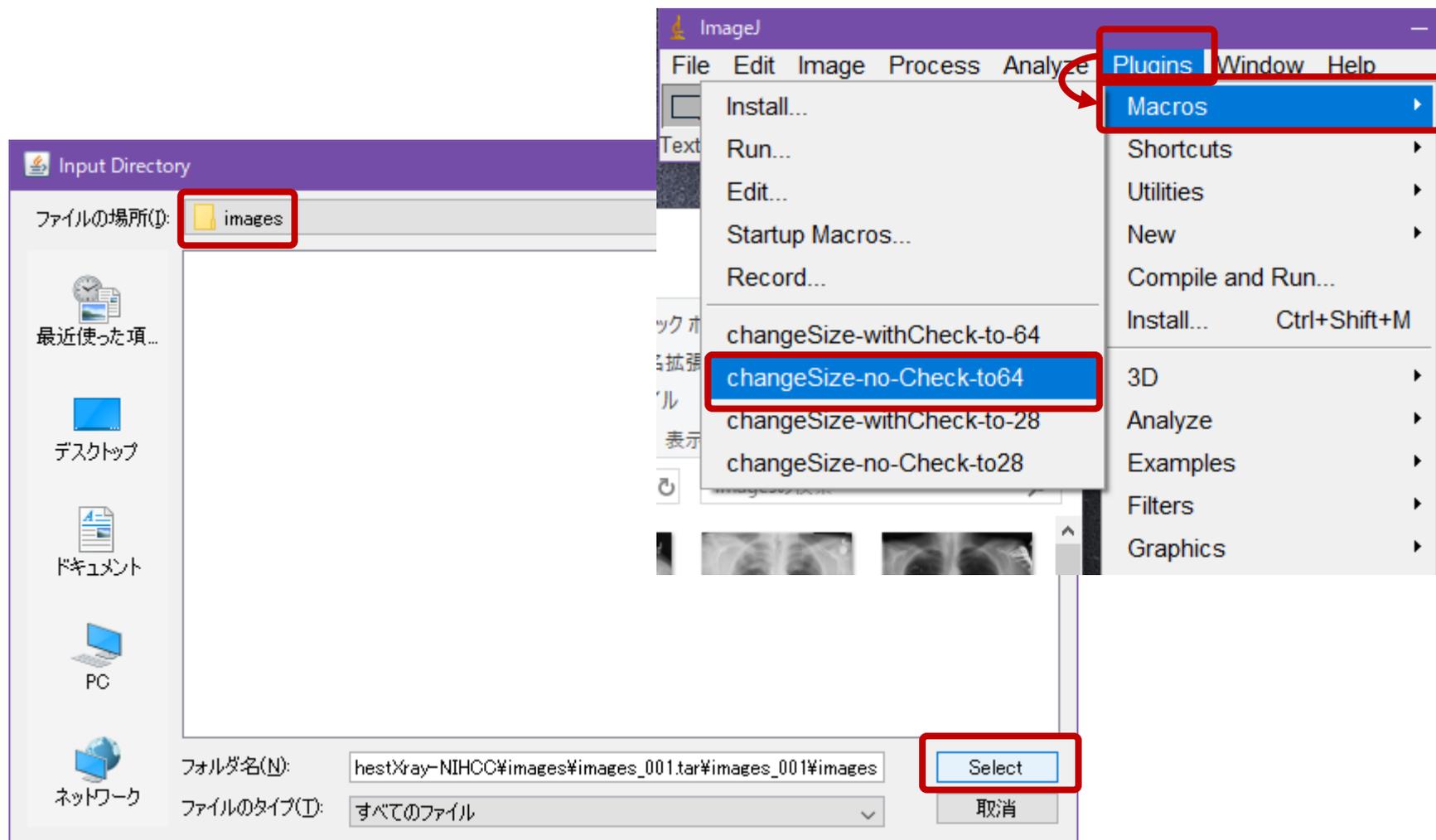
- 1学期の初回課題に並列して記載しています。
  - <https://www5.dent.niigata-u.ac.jp/~nisiyama/grad/19-01-install.pdf>
- NIHのオリジナルのImageJでなくFijiを利用する場合、下記から適切なものをダウンロードしてインストールしてください。
  - <https://fiji.sc/>

## [4] 64x64pixel画像、 および28x28pixel画像への変換

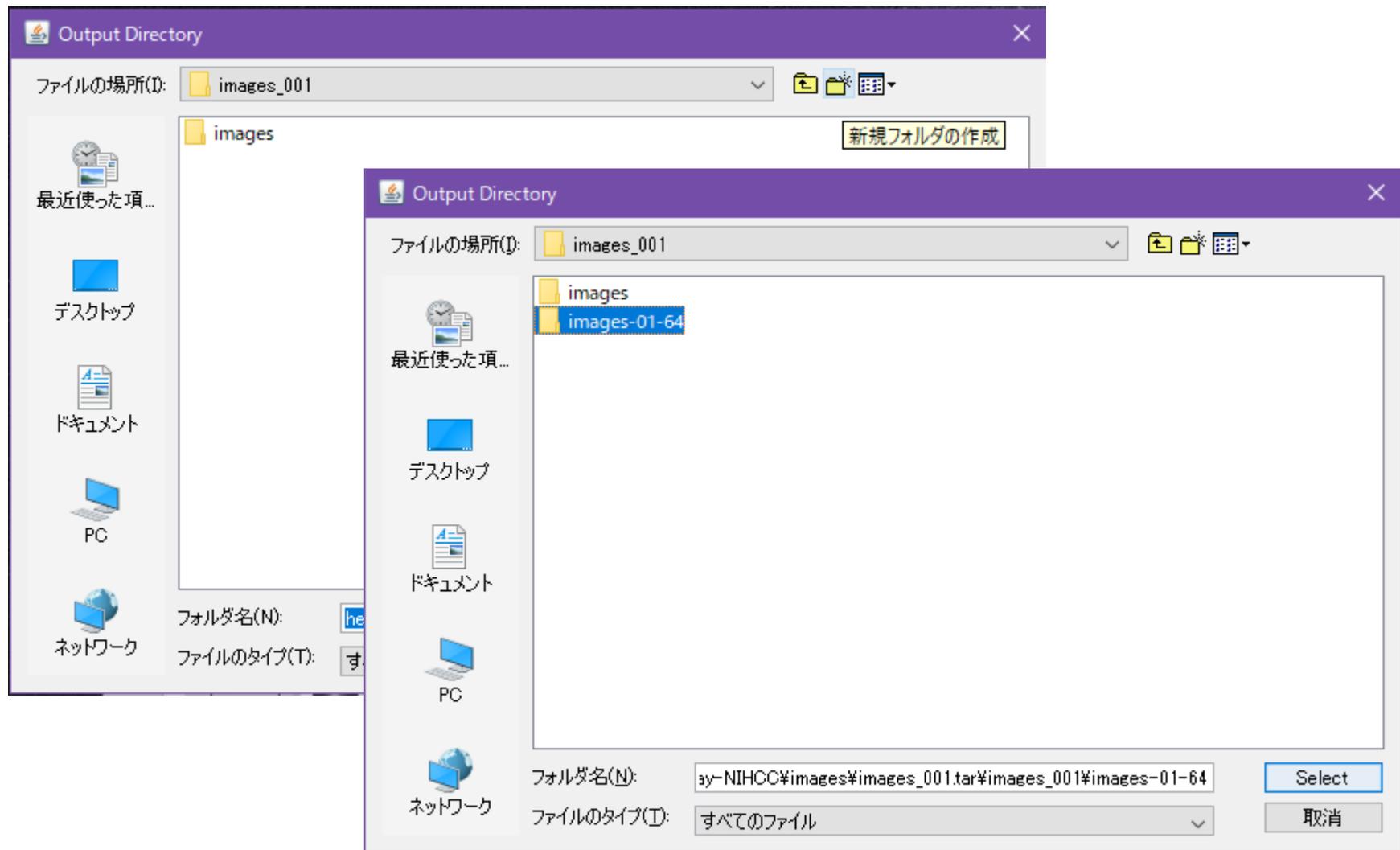
- マクロファイル(ZIP圧縮)を下記からダウンロードし、解凍して「changeSize.ijm」というマクロファイルを生成し、ImageJないしFijiのマクロにインストール
- <https://www5.dent.niigata-u.ac.jp/~nisiyama/grad/changeSize.zip>



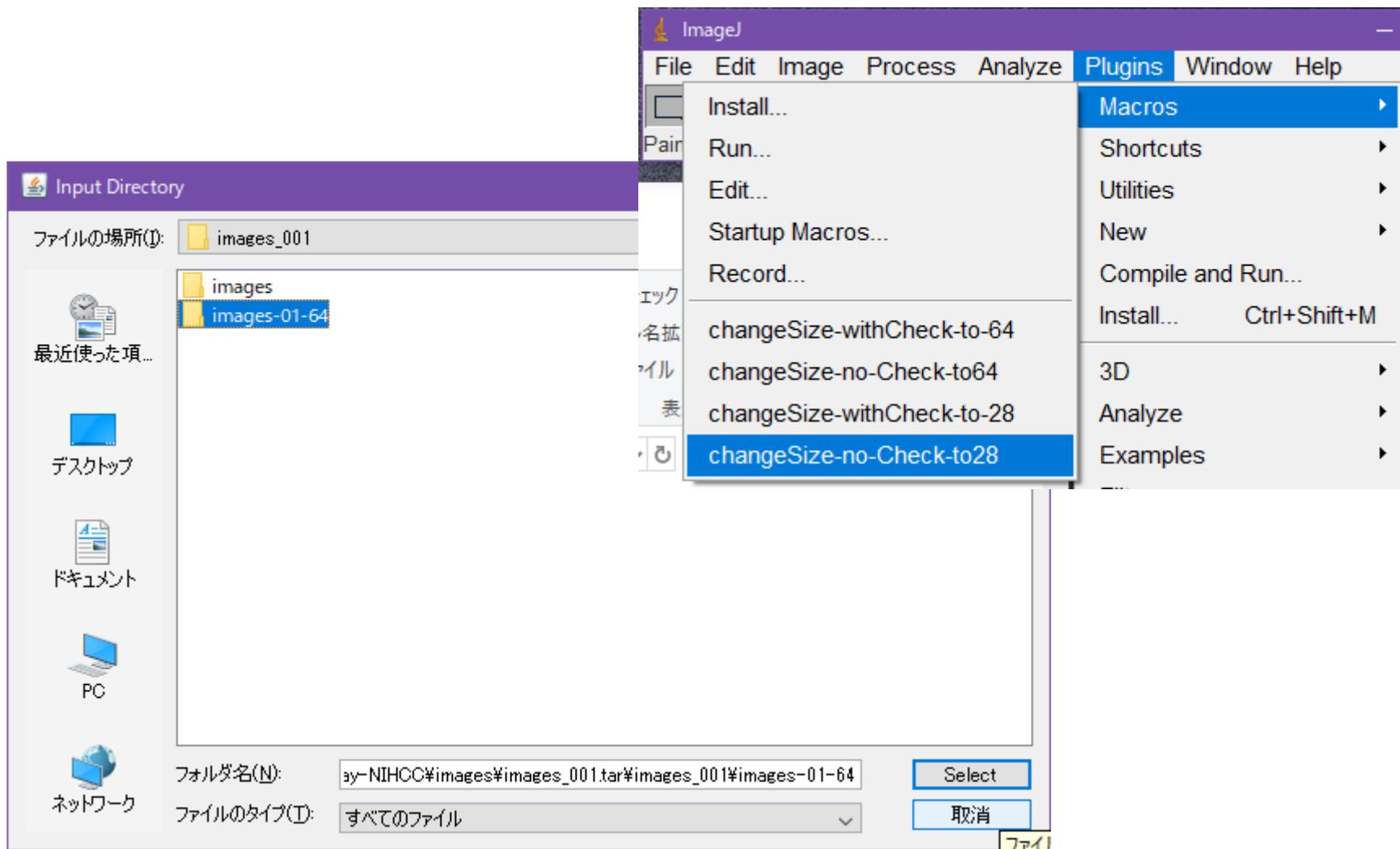
[5] マクロで「changeSize-no-Check-to64」を選択し、「Input Directory」にて、該当する1024x1024ピクセル画像の入っているフォルダを選択



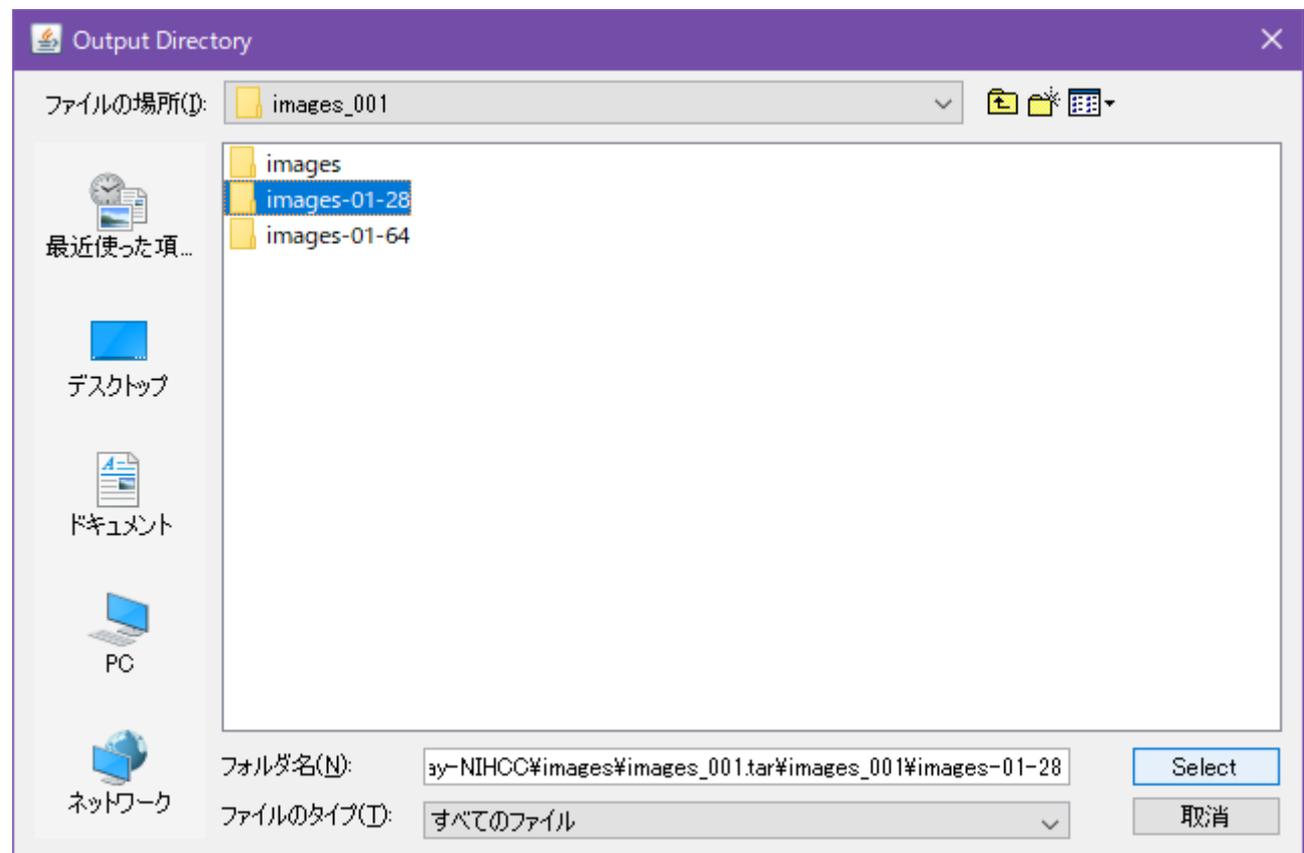
[6] 「Output Directory」にて、新規フォルダ作成し、「images-##-64」等の分かりやすい名前を付け選択



[7] マクロで「changeSize-no-Check-to28」を選択し、「Input Directory」にて、該当する64x64ピクセル画像の入っているフォルダを選択



[8] 「Output Directory」にて、新規フォルダ作成し、「images-##-28」等の分かりやすい名前を付け選択



# [9] 出来上がったファイル

← → ↑ ↓ << images >> images\_001.tar > images\_001 > images-01-64

images-01-64の検索

- ChestXray-NIHCC
  - images
    - extracted
    - images\_001.tar
    - images\_001
      - images
      - images-01-28
      - images-01-64
    - changeSize.zip
    - test

12 個の項目 | 1 個の項目を選択 3.96 KB

00000001\_000.png  
PNG ファイル

撮影日時: 撮影日の指定  
大きさ: 64 x 64

← → ↑ ↓ << images >> images\_001.tar > images\_001 > images-01-28

images-01-28の検索

- ChestXray-NIHCC
  - images
    - extracted
    - images\_001.tar
    - images\_001
      - images
      - images-01-28
      - images-01-64
    - changeSize.zip
    - test

12 個の項目 | 1 個の項目を選択 880 バイト

00000001\_000.png  
PNG ファイル

撮影日時: 撮影日の指定  
大きさ: 28 x 28

# [10] ファイルの変換後、 下記のファイル、フォルダは消去可能

- [images\_###.tar]フォルダとその中身。
- images\_###.tar.gz ファイル。

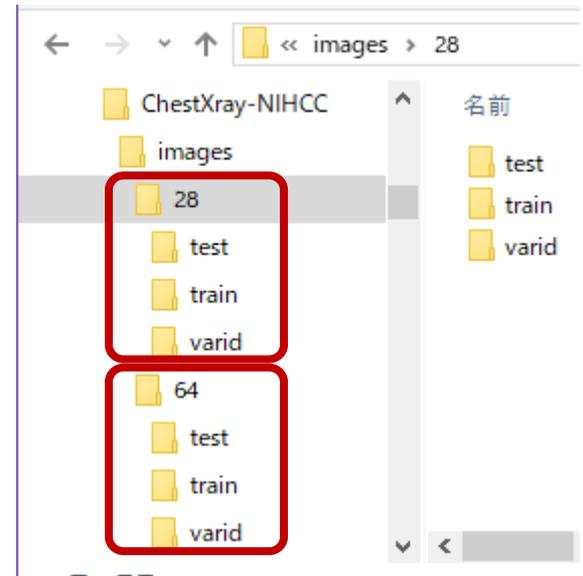
※このページまで来たら、「###」の番号を「012」まで一つずつ増やし、  
[1]に戻って次の「.tar.gz」ファイルを読み込んで、処理していく

2019.12.27  
補足追加

# フォルダをまとめる

## 28ピクセルと64ピクセルは別に行う

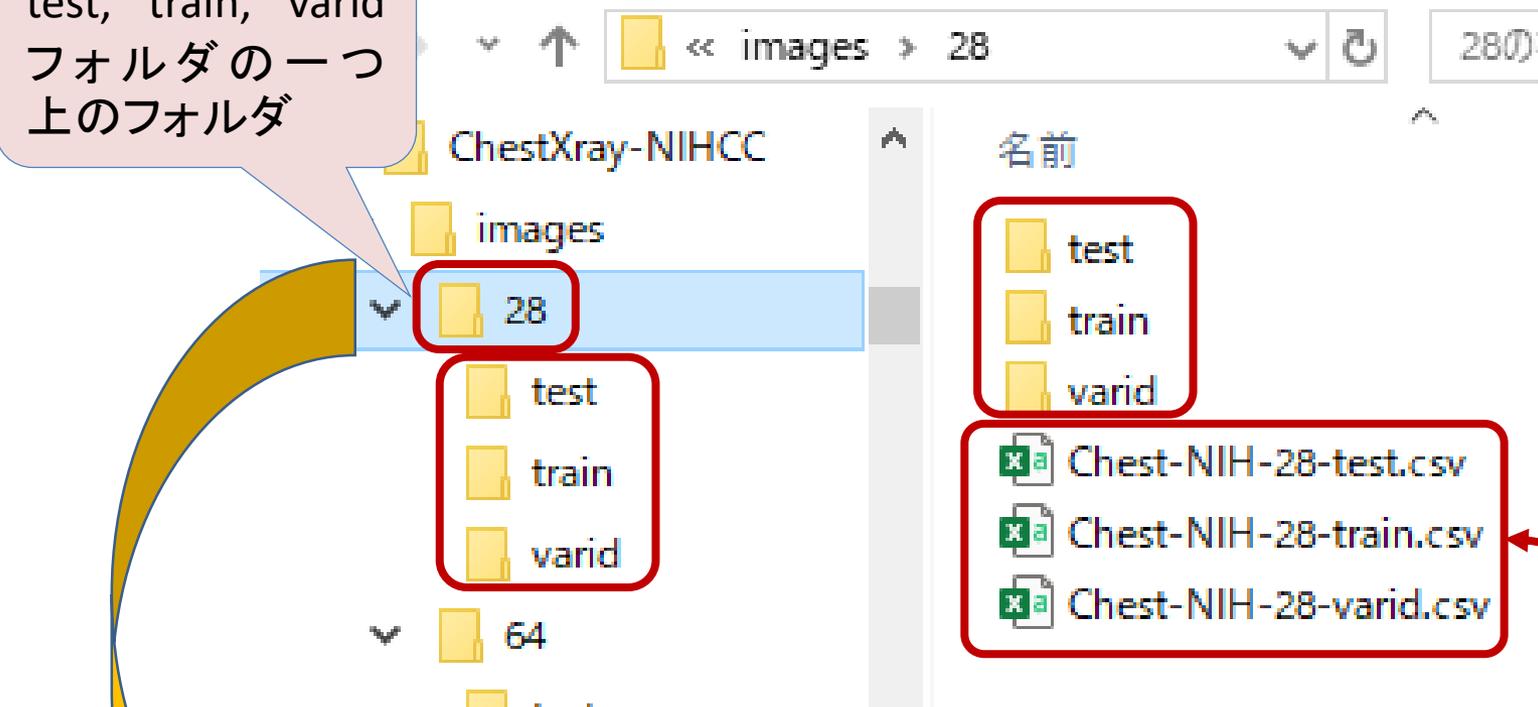
- 28ピクセル(ないし64ピクセル)の画像を下記のごとく、3つのフォルダに分ける。(画像は任意)
- test フォルダ
  - 200枚
- varid フォルダ
  - 20,000枚
- train フォルダ
  - 残り(91,920枚) —— テキストでは「92,120枚」とあるが、test画像に200枚使うと、残りは「91,920枚」となる。



# データリスト(csv)ファイルを作る

- ウィンドウズの「コマンドプロンプト」を起動させる。
- 「train, varid, test」の3つのフォルダが入っている一つ上のフォルダに移動する。
- 下記のコマンドで、各ディレクトリのファイル名一覧を「CSV」ファイルとして保存する。(ファイル名には28ピクセル用として28という数字を入れているが、64ピクセルの場合には64を入れるとわかりやすい)
  - dir /b train > Ches-NIH-28-train.csv
  - dir /b varid > Ches-NIH-28-varid.csv
  - dir /b train > Ches-NIH-28-train.csv

test, train, varid  
フォルダの一つ  
上のフォルダ



```
C:\Users\HN\Downloads\ChestXray-NIHCC\images\28>dir /b train > Chest-NIH-28-train.csv
C:\Users\HN\Downloads\ChestXray-NIHCC\images\28>dir /b varid > Chest-NIH-28-varid.csv
C:\Users\HN\Downloads\ChestXray-NIHCC\images\28>dir /b test > Chest-NIH-28-test.csv
C:\Users\HN\Downloads\ChestXray-NIHCC\images\28>
```

コマンドプロンプト内で、CD コマンドにて、該当するディレクトリに  
移動後に、実施すること

Ex. `cd Downloads\ChestXray-NIHCC\images\28`

以下[2-1]から[2-7]までtest, varid, trainの  
3つのcsvファイルで繰り返して操作する

文中の「test」および「28」、さらにセルの最下行の番号等は、各データ・ファイル毎に異なるので注意！

A1

csvファイルを開いた直後、セル「A1」を選択状態。  
【End】キーを押し、【↓】を押して、A列のデータの一番下のセルに移動

	A	B	C	D	E	F
1	00000001_000.png					
2	00000001_001.png					
3	00000001_002.png					
4	00000002_000.png					
5	00000003_000.png					
6	00000003_001.png					
7	00000003_002.png					
8	00000003_003.png					
9	00000003_004.png					
10	00000003_005.png					
11	00000003_006.png					
12	00000003_007.png					
13	00000004_000.png					
14	00000005_000.png					
15	00000005_001.png					
16	00000005_002.png					
17	00000005_003.png					
18	00000005_004.png					
19	00000005_005.png					
20	00000005_006.png					

	A	B	C
182	00000037_000.png		
183	00000038_000.png		
184	00000038_001.png		
185	00000038_002.png		
186	00000038_003.png		
187	00000038_004.png		
188	00000038_005.png		
189	00000038_006.png		
190	00000038_007.png		
191	00000039_000.png		
192	00000039_001.png		
193	00000039_002.png		
194	00000039_003.png		
195	00000039_004.png		
196	00000040_000.png		
197	00000040_001.png		
198	00000040_002.png		
199	00000040_003.png		
200	00000041_000.png		

対象が「test」の場合であり、対象が「varid」、「train」の場合にはそれぞれの名前(フォルダ名)にすること

A列最下行の右隣りのB列のセルに移動し

=“.¥test¥” &

まで、入力し、【←】にて左隣のセル(A列)を選択して、最下行のセル番地を入力。その後【Enter】

※今回の場合

=“.¥test¥” & A200

となる。

最後のデータの列記号・行番号

※セル番地を手入力しても良い

※「test」は、対象が「test」の時に入力する文字(ディレクトリ名)で、対象が「train」や「varid」の場合には、それぞれの文字(対象となるディレクトリ名)にすること。

今回は、最後のデータの行番号が200の場合

183	00000038_000.png
184	00000038_001.png
185	00000038_002.png
186	00000038_003.png
187	00000038_004.png
188	00000038_005.png
189	00000038_006.png
190	00000038_007.png
191	00000039_000.png
192	00000039_001.png
193	00000039_002.png
194	
195	
196	
197	
198	00000040_002.png
199	00000040_003.png
200	00000041_000.png

200	00000041_000.png	=“.¥test¥”&A200
-----	------------------	-----------------

200	00000041_000.png	.¥test¥00000041_000.png
201		
202		

B199

	A	B
183	00000038_000.png	
184	00000038_001.png	
185	00000038_002.png	
186	00000038_003.png	
187	00000038_004.png	
188	00000038_005.png	
189	00000038_006.png	
190	00000038_007.png	
191	00000039_000.png	
192	00000039_001.png	
193	00000039_002.png	
194	00000039_003.png	
195	00000039_004.png	
196	0	
197	0	
198	0	
199	00000040_003.png	
200	00000041_000.png	=test#00000041_000.png
201		

②一つ上のセルを選択(クリック)  
(数式の貼り付け先最下行)

数式の  
コピー元

①計算されたB列の最下行のセルを選択し、【ctrl】+【c】等でコピー元とする。

③【End】キーを押した後、【Shift】+【↑】にて、B列最下行の一つ上からトップのセルまでを選択。

	A	B	C
1	00000001_000.png		
2	00000001_001.png		
3	00000001_002.png		
4	00000002_000.png		
5	00000002_001.png		
6	00000002_002.png		
7	00000003_002.png		
8	00000003_003.png		
9	00000003_004.png		
10	00000003_005.png		
11	00000003_006.png		
12	00000003_007.png		
13	00000004_000.png		
14	00000005_000.png		
15	00000005_001.png		
16	00000005_002.png		
17	00000005_003.png		
18	00000005_004.png		
19	00000005_005.png		
20	00000005_006.png		

数式の貼り付け先最上行

Chest-NIH-28-test

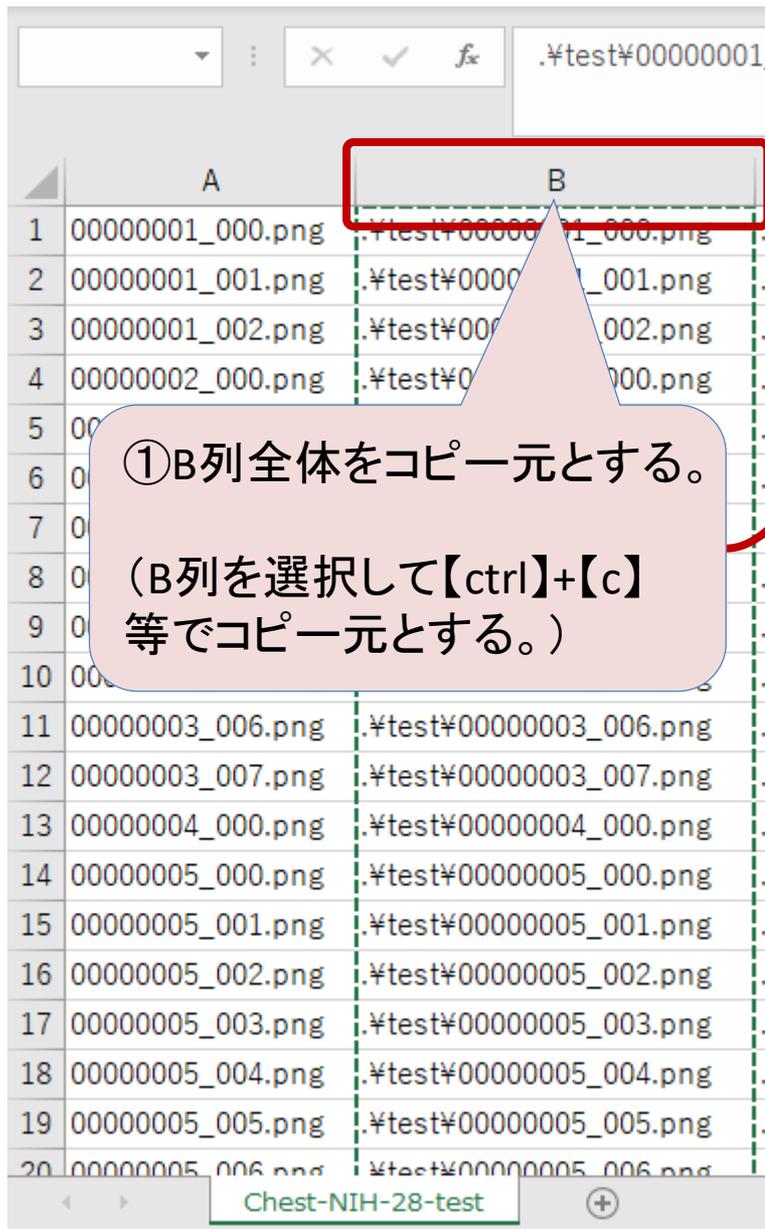
B199 : x ✓ fx =".¥test¥"&A199

	A	B
190	00000038_007.png	.¥test¥00000038_007.png
191	00000039_000.png	.¥test¥00000039_000.png
192	00000039_001.png	.¥test¥00000039_001.png
193	00000039_002.png	.¥test¥00000039_002.png
194	00000039_003.png	.¥test¥00000039_003.png
195	00000039_004.png	.¥test¥00000039_004.png
196	00000040_000.png	.¥test¥00000040_000.png
197	00000040_001.png	.¥test¥00000040_001.png
198	00000040_002.png	.¥test¥00000040_002.png
199	00000040_003.png	.¥test¥00000040_003.png
200	00000041_000.png	.¥test¥00000041_000.png
201		
202		
203		
204		
205		
206		
207		
208		
209		

Chest-NIH-28-test (+)

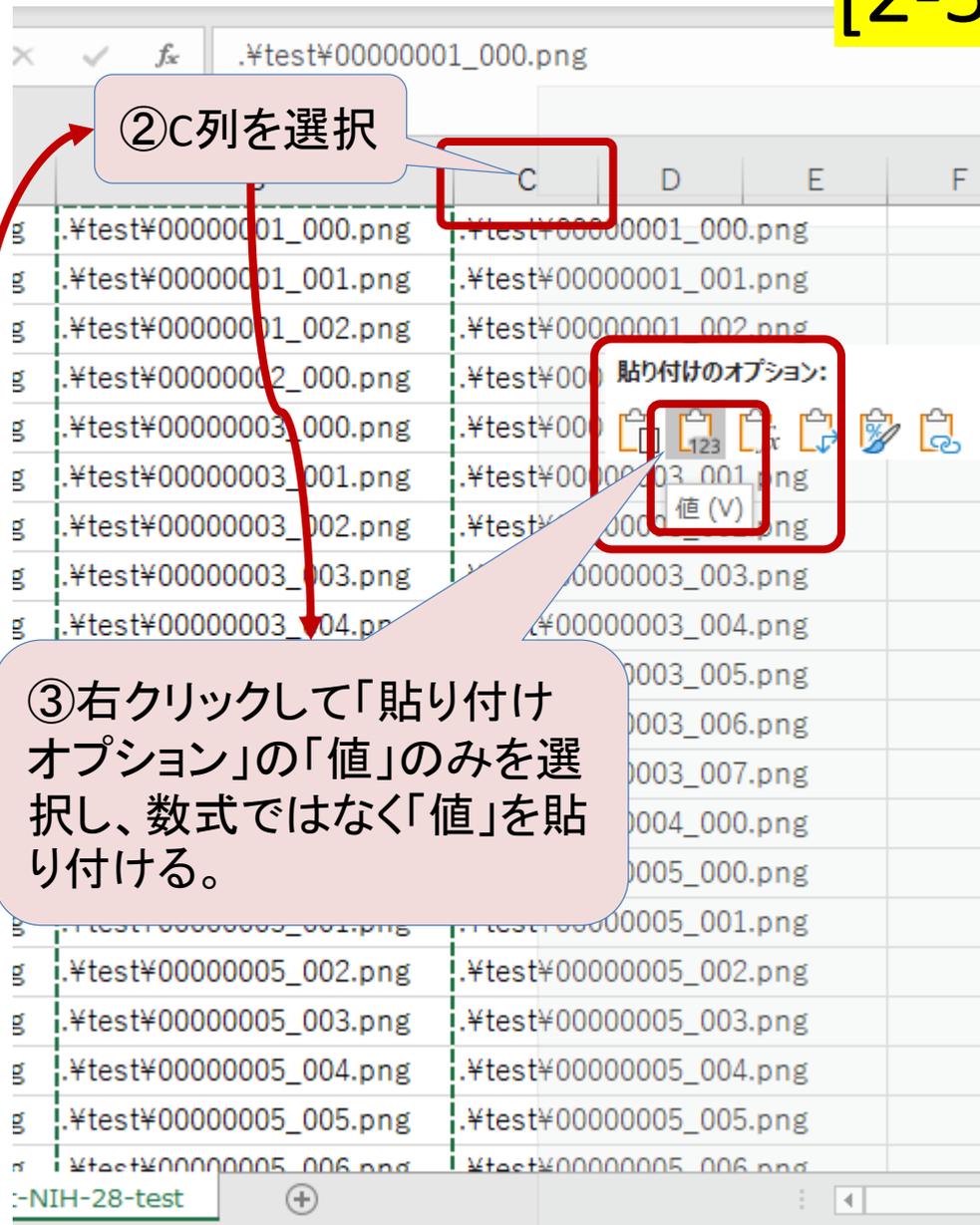
コピー先を選択し、Enter キーを押すか、貼り付けを選択します。

【Enter】を押して、B列最下行の式を、  
B列の1行目までペースト。



①B列全体をコピー元とする。  
(B列を選択して【ctrl】+【c】  
等でコピー元とする。)

	A	B
1	00000001_000.png	¥test¥00000001_000.png
2	00000001_001.png	¥test¥00000001_001.png
3	00000001_002.png	¥test¥00000001_002.png
4	00000002_000.png	¥test¥00000002_000.png
5	00000002_001.png	¥test¥00000002_001.png
6	00000002_002.png	¥test¥00000002_002.png
7	00000003_000.png	¥test¥00000003_000.png
8	00000003_001.png	¥test¥00000003_001.png
9	00000003_002.png	¥test¥00000003_002.png
10	00000003_003.png	¥test¥00000003_003.png
11	00000003_006.png	¥test¥00000003_006.png
12	00000003_007.png	¥test¥00000003_007.png
13	00000004_000.png	¥test¥00000004_000.png
14	00000005_000.png	¥test¥00000005_000.png
15	00000005_001.png	¥test¥00000005_001.png
16	00000005_002.png	¥test¥00000005_002.png
17	00000005_003.png	¥test¥00000005_003.png
18	00000005_004.png	¥test¥00000005_004.png
19	00000005_005.png	¥test¥00000005_005.png
20	00000005_006.png	¥test¥00000005_006.png



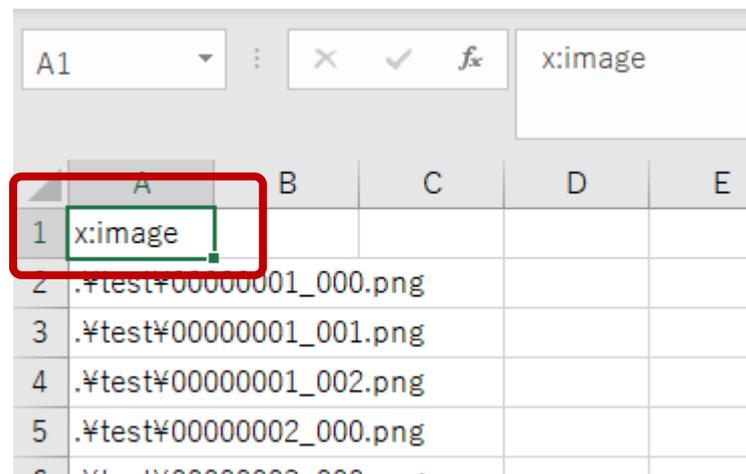
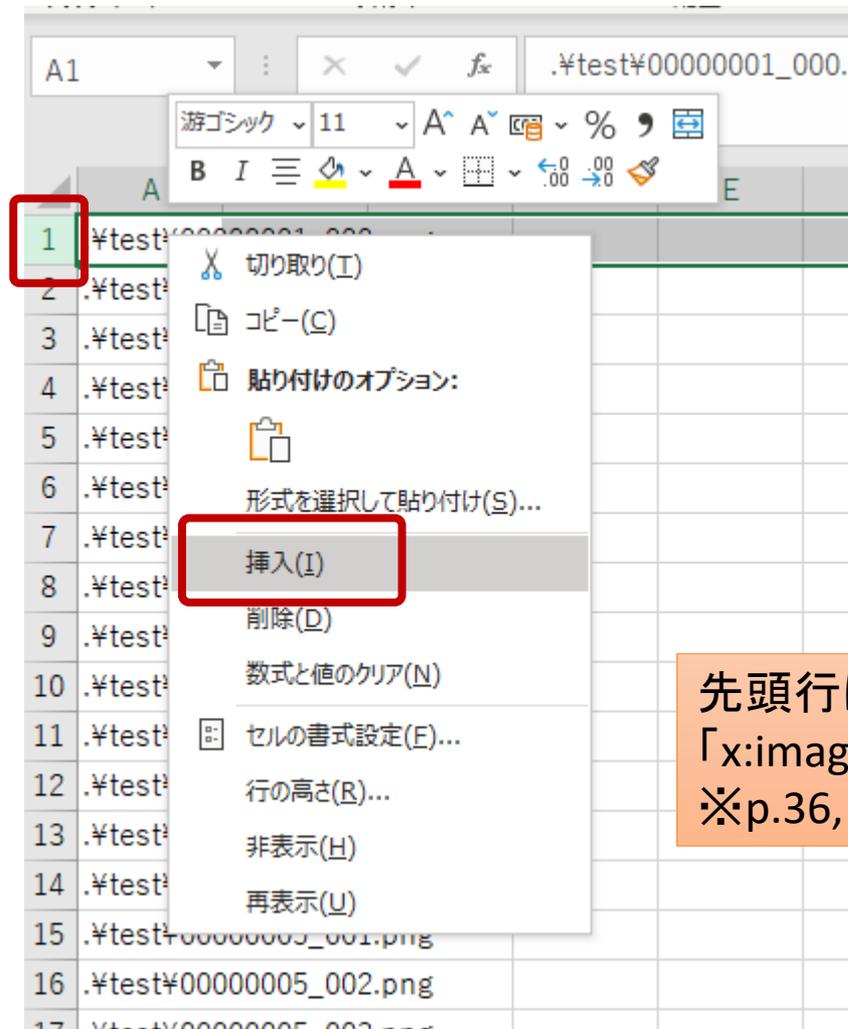
②C列を選択

③右クリックして「貼り付けオプション」の「値」のみを選択し、数式ではなく「値」を貼り付ける。

貼り付けのオプション:  
値 (V)

	C	D	E	F
1	g ¥test¥00000001_000.png	¥test¥00000001_000.png		
2	g ¥test¥00000001_001.png	¥test¥00000001_001.png		
3	g ¥test¥00000001_002.png	¥test¥00000001_002.png		
4	g ¥test¥00000002_000.png	¥test¥00000002_000.png		
5	g ¥test¥00000003_000.png	¥test¥00000003_000.png		
6	g ¥test¥00000003_001.png	¥test¥00000003_001.png		
7	g ¥test¥00000003_002.png	¥test¥00000003_002.png		
8	g ¥test¥00000003_003.png	¥test¥00000003_003.png		
9	g ¥test¥00000003_004.png	¥test¥00000003_004.png		
10	g ¥test¥00000003_005.png	¥test¥00000003_005.png		
11	g ¥test¥00000003_006.png	¥test¥00000003_006.png		
12	g ¥test¥00000003_007.png	¥test¥00000003_007.png		
13	g ¥test¥00000004_000.png	¥test¥00000004_000.png		
14	g ¥test¥00000005_000.png	¥test¥00000005_000.png		
15	g ¥test¥00000005_001.png	¥test¥00000005_001.png		
16	g ¥test¥00000005_002.png	¥test¥00000005_002.png		
17	g ¥test¥00000005_003.png	¥test¥00000005_003.png		
18	g ¥test¥00000005_004.png	¥test¥00000005_004.png		
19	g ¥test¥00000005_005.png	¥test¥00000005_005.png		
20	g ¥test¥00000005_006.png	¥test¥00000005_006.png		





先頭行に1行挿入し  
「x:image」といった文字を入力しておく。  
※p.36, p.66等を参照のこと

[2-1]から、このページ([2-7])まで、「test」フォルダ(csvは28ピクセル・test用のファイル)に対して行った操作を、varidおよびtrainの各フォルダと対応する各csvファイルに対し行うこと。また、64ピクセルの各フォルダ、csvファイルに対しても行うこと。

2.10.2 (p.72-74)で変更・修正した  
mnist\_dcgan\_with\_lavel.sdcproj※  
に、該当するDATASETを割り当てていく

- ※プロジェクトにブロックを挿入したり削除したり、あるいは接続や各ブロックのパラメータ等を変更・修正する前に、元プロジェクトを残す目的で、複製し別名で保存していくこと。
- DATASETへの割り当てについては、これまでのテキストの記載内容を参照すること。

# GPU非搭載の場合

- GPU非搭載の場合、計算時間が長くなりすぎるので、エポック数を5位からスタートしてください。  
※テキストではバッチサイズ・エポック数はそのままと記載ありますが、GPU搭載の場合でも、エポック数は減らした状態からチェックした方が良いと思います。
- また、trainデータの数を減らした状態（半分とか1/4とか）での影響等を確認してください。
- データ数削減の影響の方が面白い結果に繋がると思います。

# データフォルダはそのまま 利用データのみ変更する方法

- テキストの「2.8.1」p.49-50にあるように、csvファイル内にて、利用するデータリストを変えるだけで、画像フォルダはそのまま利用できます。
- たとえば、trainデータを25000までに減らしたい場合、csvファイルの25002行目以降を削除して、タイトル行1行＋データ行25000とします。
- ファイル名を「Chest-NIH-28-train-25000.csv」と名前を変更して保存し、プロジェクトのデータセットにてtrainデータセットの対象を上記のcsvファイルに付け替えるとOKです。