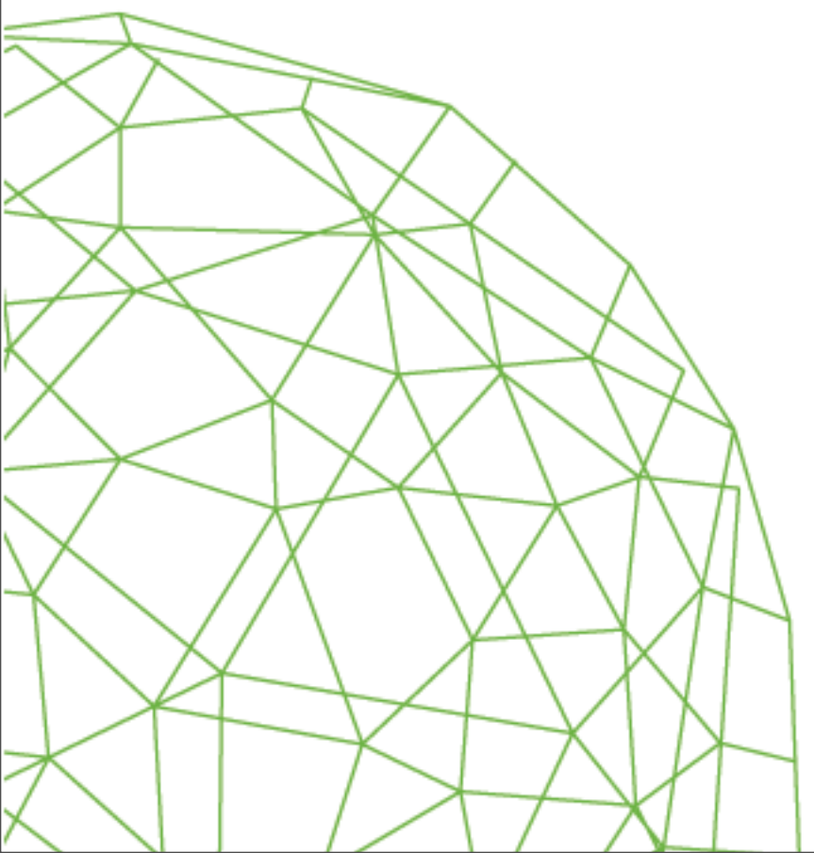
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

CMM
Center for
Mathematical
Modeling

Demo: obtaining cutouts from SDSS



Guillermo Cabrera
CMM, Universidad de Chile

Obtaining Cutouts from SDSS

Obtaining Cutouts from SDSS

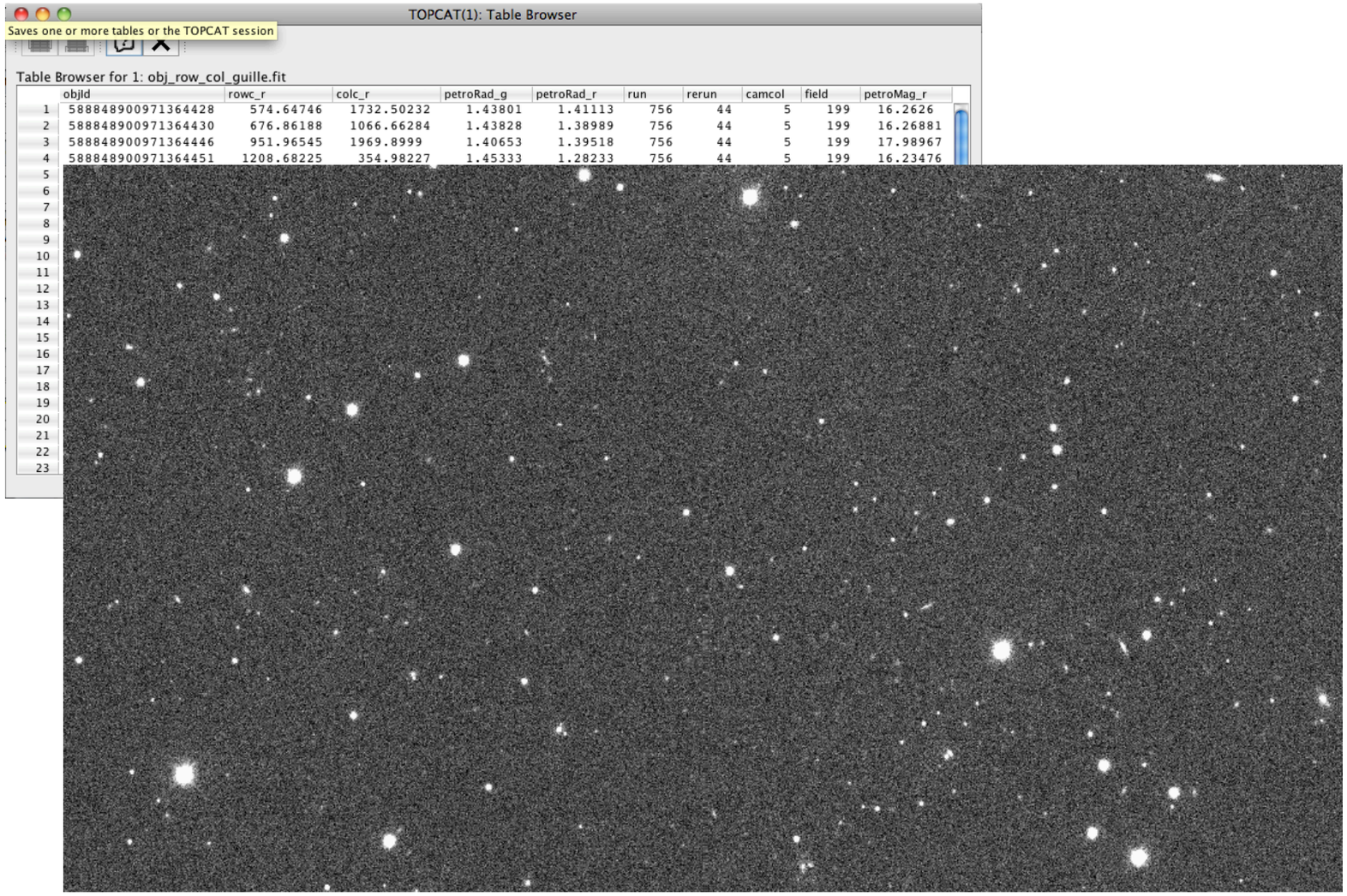
TOPCAT(1): Table Browser

Saves one or more tables or the TOPCAT session

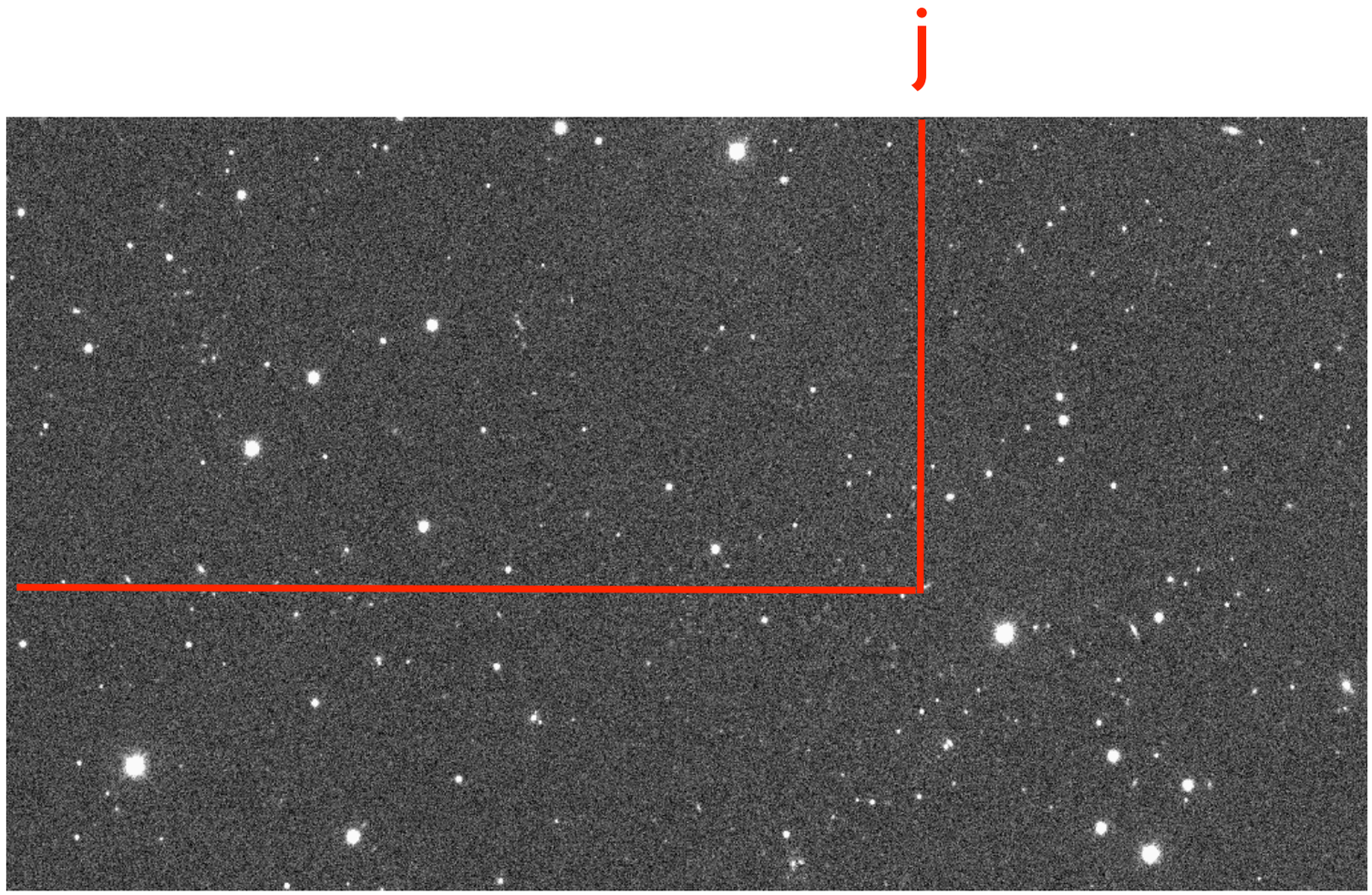
Table Browser for 1: obj_row_col_guille.fit

	objId	rowc_r	colc_r	petroRad_g	petroRad_r	run	rerun	camcol	field	petroMag_r
1	588848900971364428	574.64746	1732.50232	1.43801	1.41113	756	44	5	199	16.2626
2	588848900971364430	676.86188	1066.66284	1.43828	1.38989	756	44	5	199	16.26881
3	588848900971364446	951.96545	1969.8999	1.40653	1.39518	756	44	5	199	17.98967
4	588848900971364451	1208.68225	354.98227	1.45333	1.28233	756	44	5	199	16.23476
5	588848900971364457	1270.63232	1098.47229	1.60994	2.19224	756	44	5	199	16.42279
6	588848900971364482	1289.84619	891.37158	1.45655	1.32003	756	44	5	199	17.3641
7	588848900971364484	1327.46729	593.16467	1.48057	1.29462	756	44	5	199	16.34192
8	588848900971364492	149.33058	1614.99695	1.41921	1.45908	756	44	5	199	16.8724
9	588848900971364505	248.26163	1172.94531	1.43005	1.43577	756	44	5	199	17.74933
10	588848900971364509	331.1709	681.07166	1.43374	1.41923	756	44	5	199	17.79399
11	588848900971364520	446.01651	465.50851	1.40243	1.36929	756	44	5	199	17.07066
12	588848900971364522	500.36752	738.21783	1.44192	1.37903	756	44	5	199	17.6296
13	588848900971364524	534.19757	26.55153	1.34153	1.3418	756	44	5	199	17.66015
14	588848900971364534	646.13995	755.28229	1.42393	1.37804	756	44	5	199	17.99287
15	588848900971364538	770.20422	997.11499	1.42436	1.37406	756	44	5	199	17.50831
16	588848900971364549	905.85846	1583.44409	1.47688	1.39276	756	44	5	199	17.31635
17	588848900971364560	1114.78882	246.49637	1.38777	1.29355	756	44	5	199	17.70358
18	588848900971364564	1153.02185	1935.00024	1.44794	1.36904	756	44	5	199	17.77866
19	588848900971364569	1182.45825	23.92278	1.4289	1.266	756	44	5	199	17.09027
20	588848900971364579	1370.04114	1574.55798	1.49369	1.36375	756	44	5	199	16.79921
21	588848900971364580	1378.06689	1573.60364	1.34412	1.2479	756	44	5	199	17.7473
22	588848900971364598	471.71838	2014.63745	8.23456	6.86085	756	44	5	199	17.75352
23	588848900971429915	261.01761	768.92902	1.49208	1.29233	756	44	5	200	16.23797

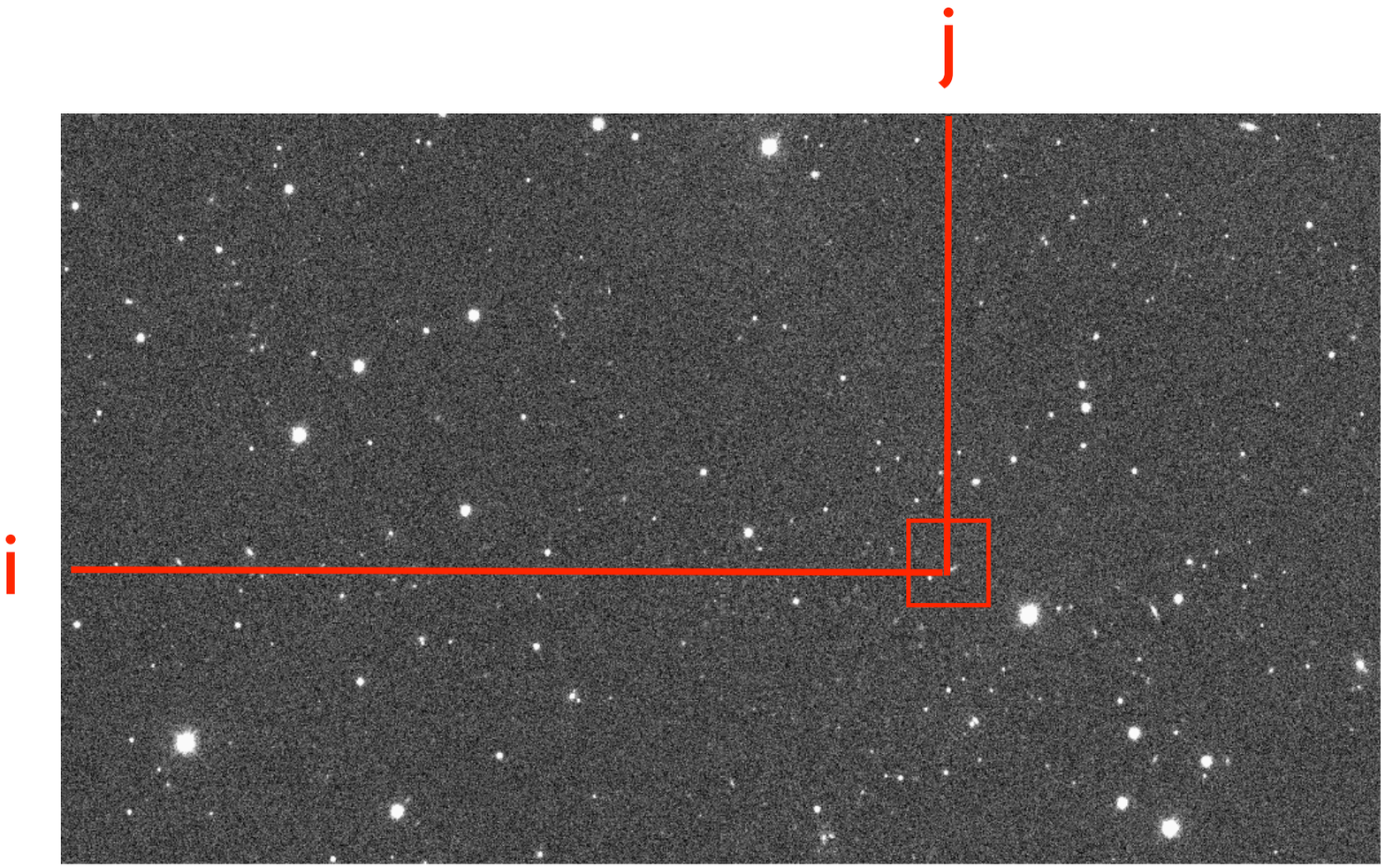
Obtaining Cutouts from SDSS



Obtaining Cutouts from SDSS



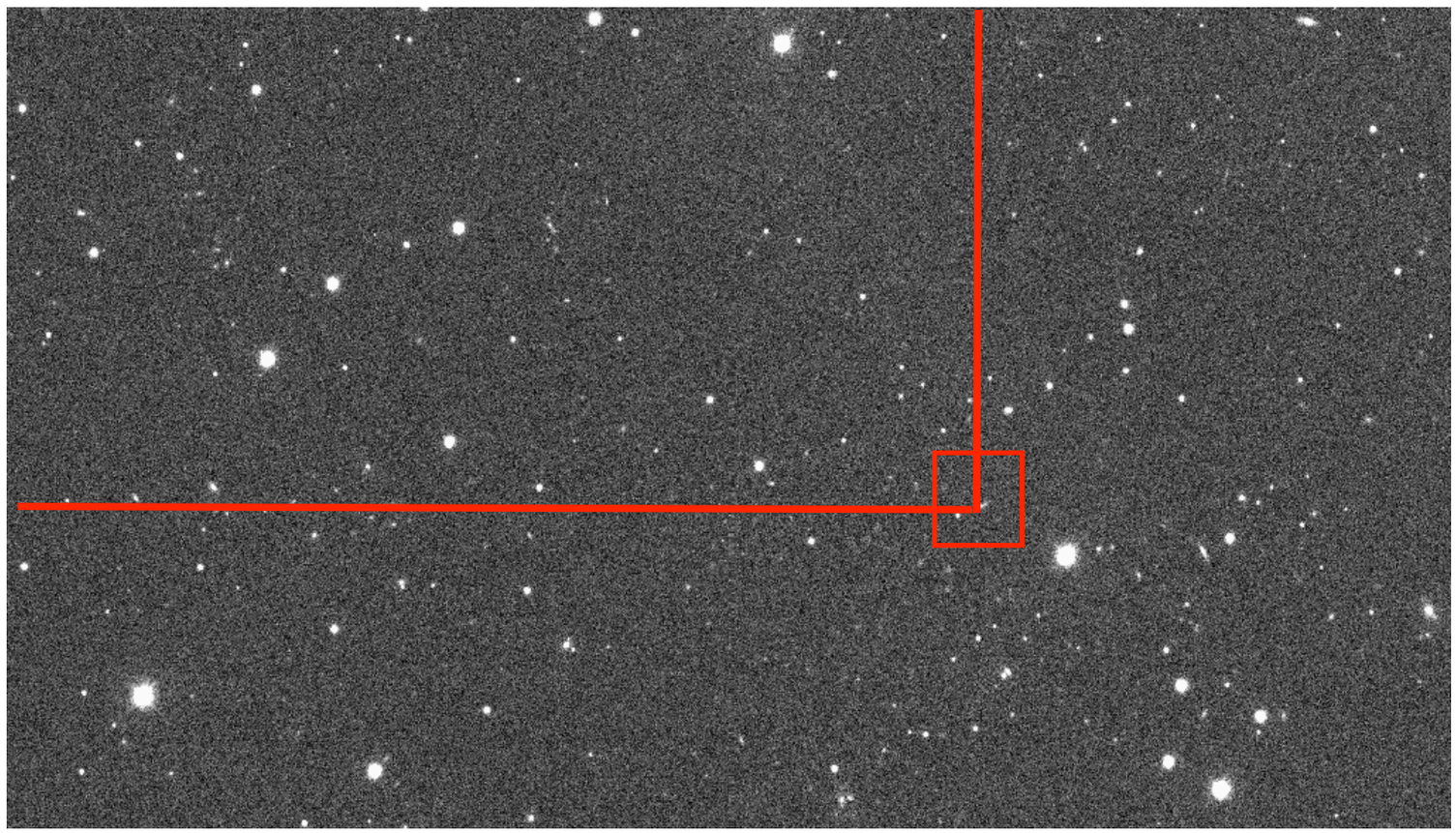
Obtaining Cutouts from SDSS



Obtaining Cutouts from SDSS

j

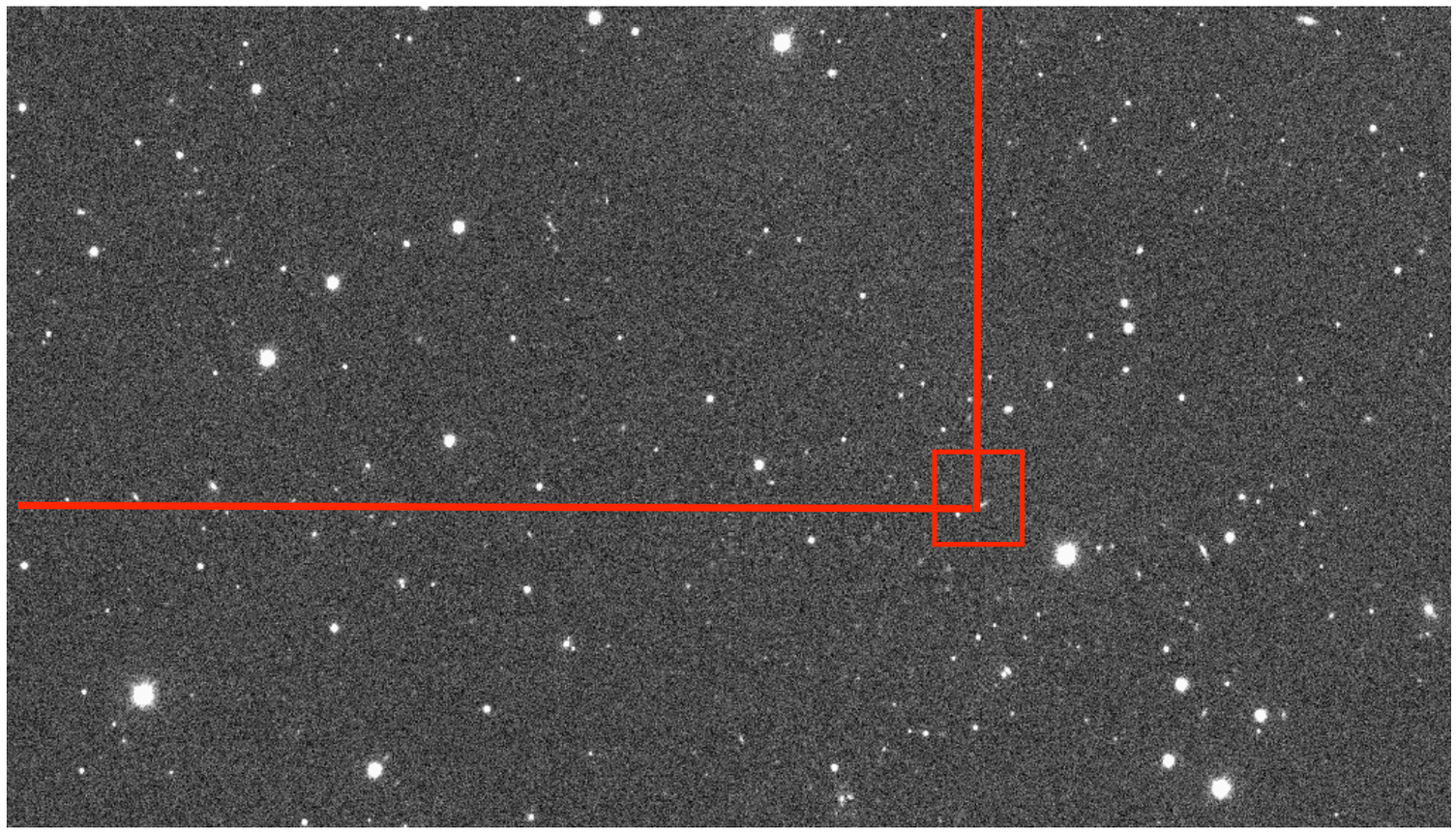
i



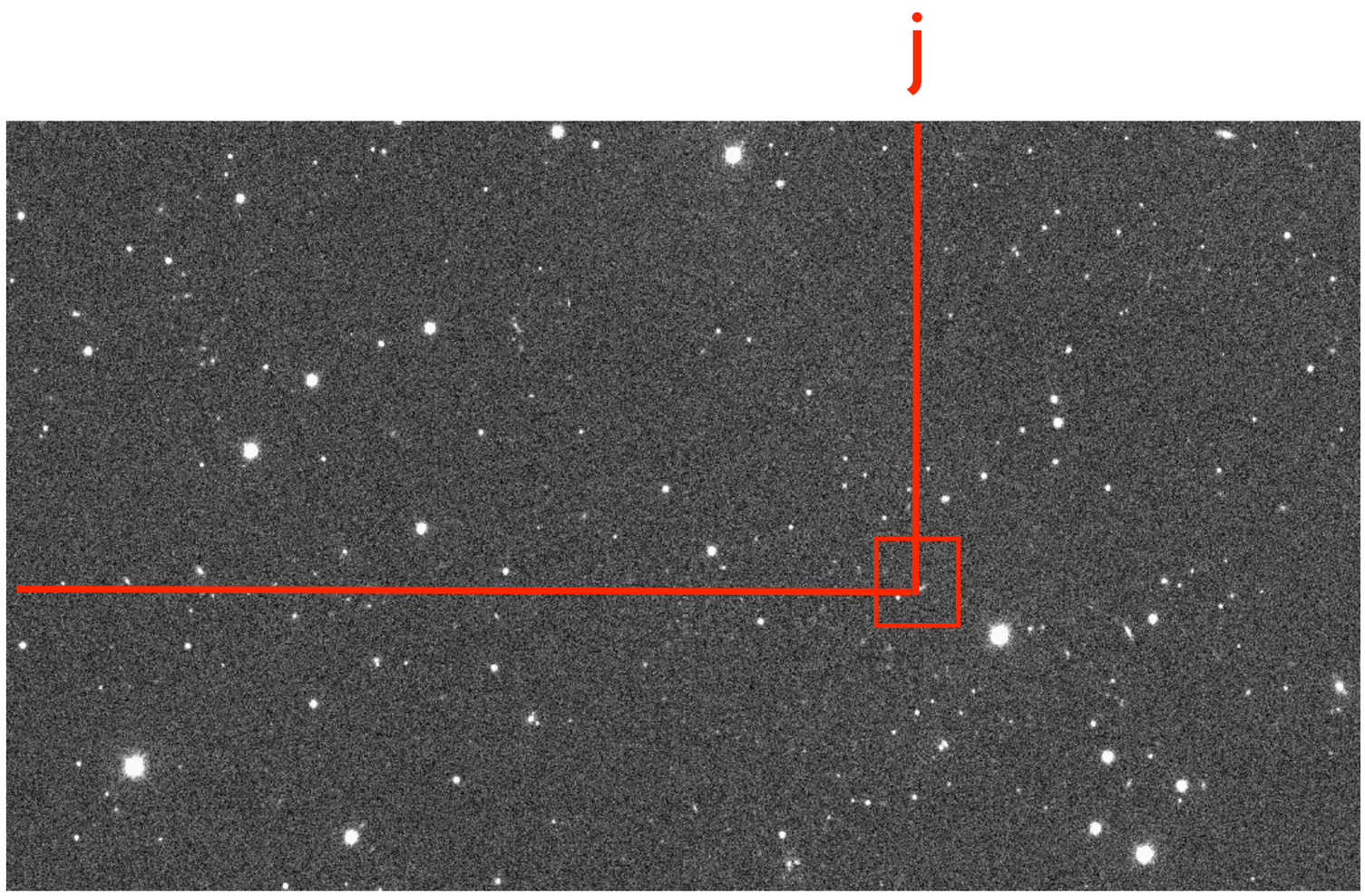
Obtaining Cutouts from SDSS

j

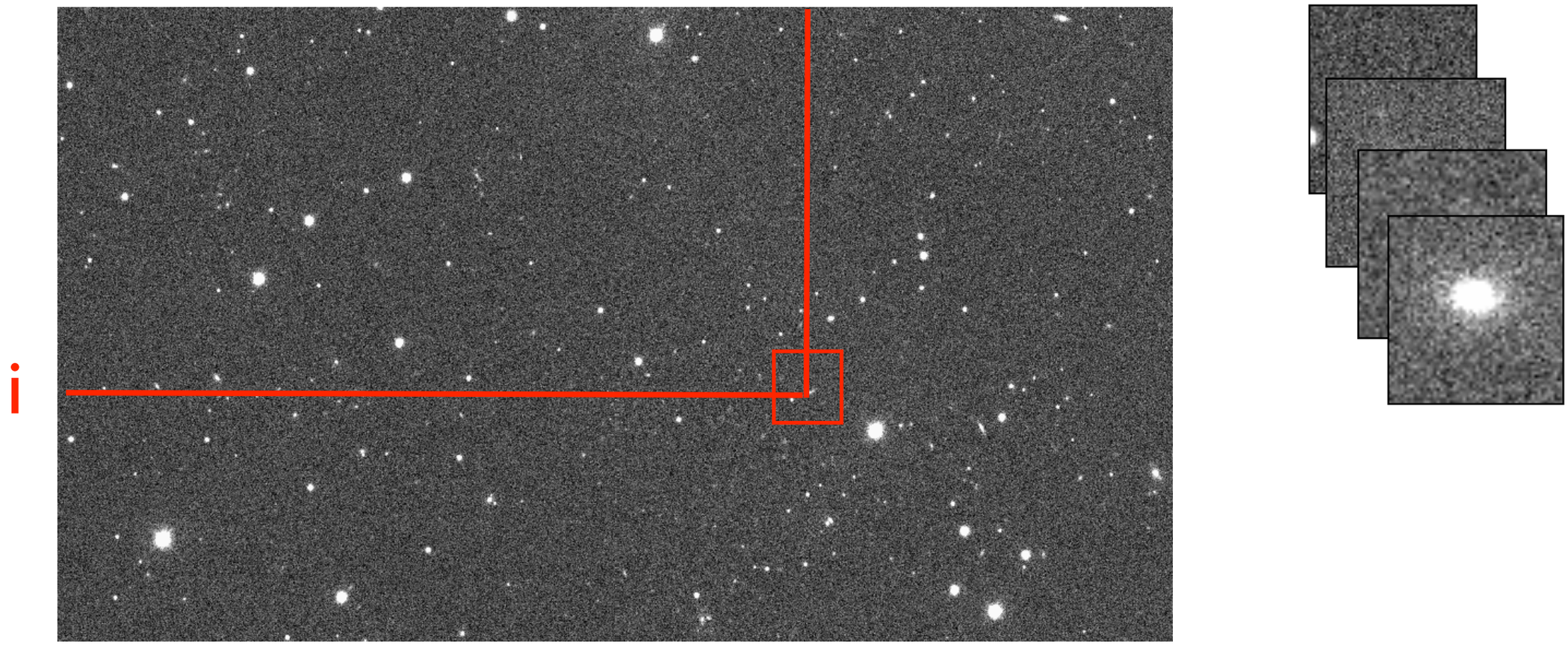
i



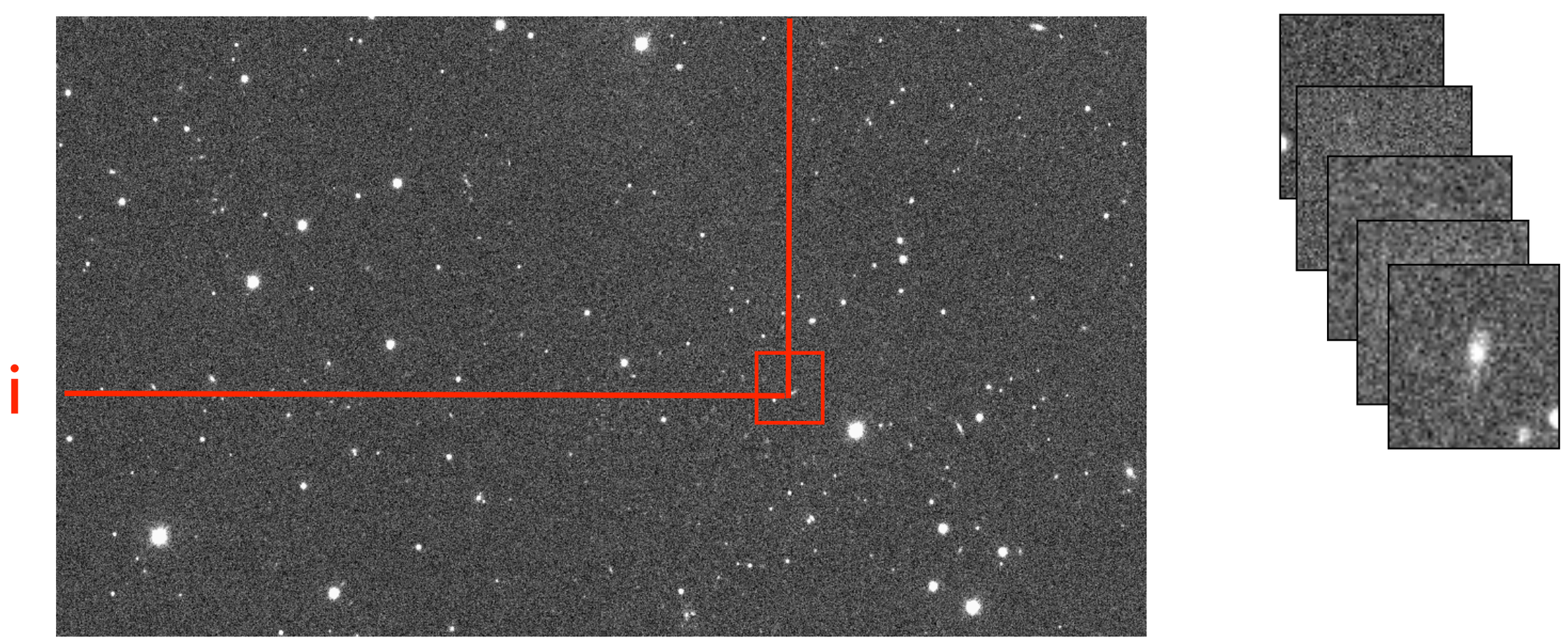
Obtaining Cutouts from SDSS



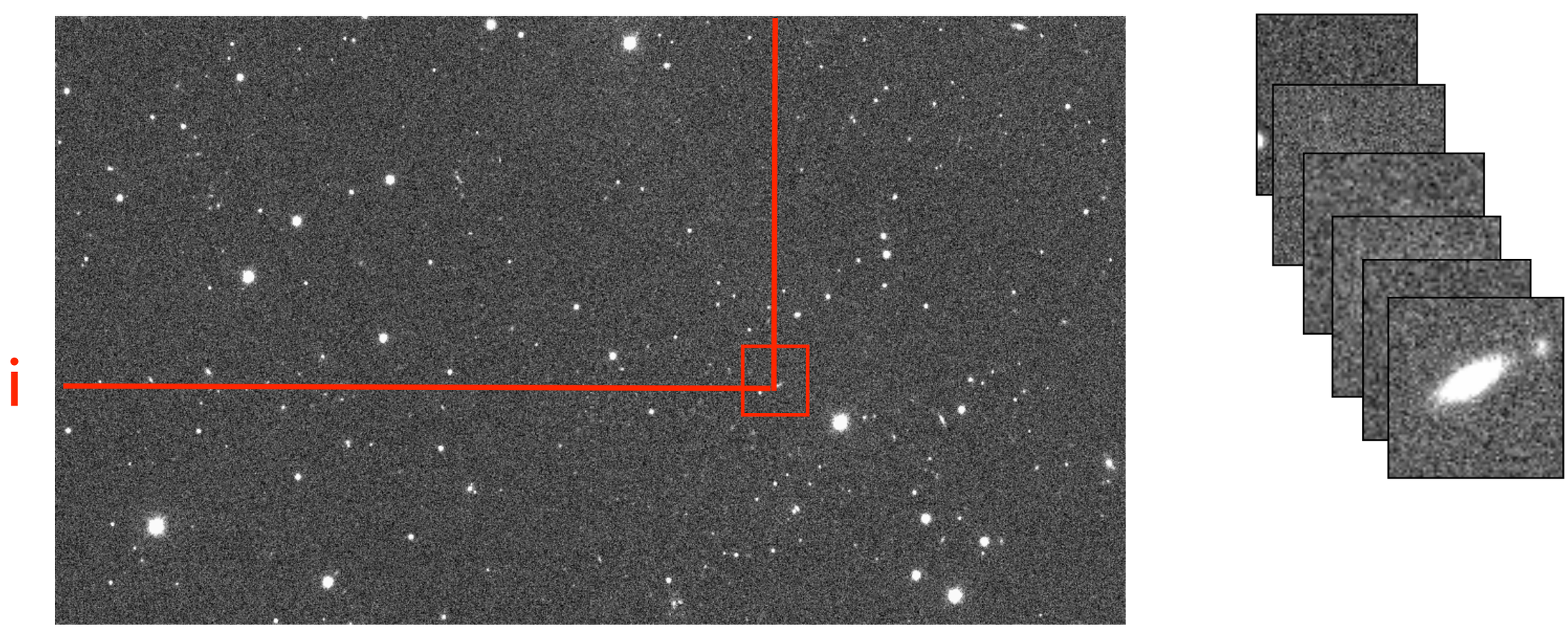
Obtaining Cutouts from SDSS



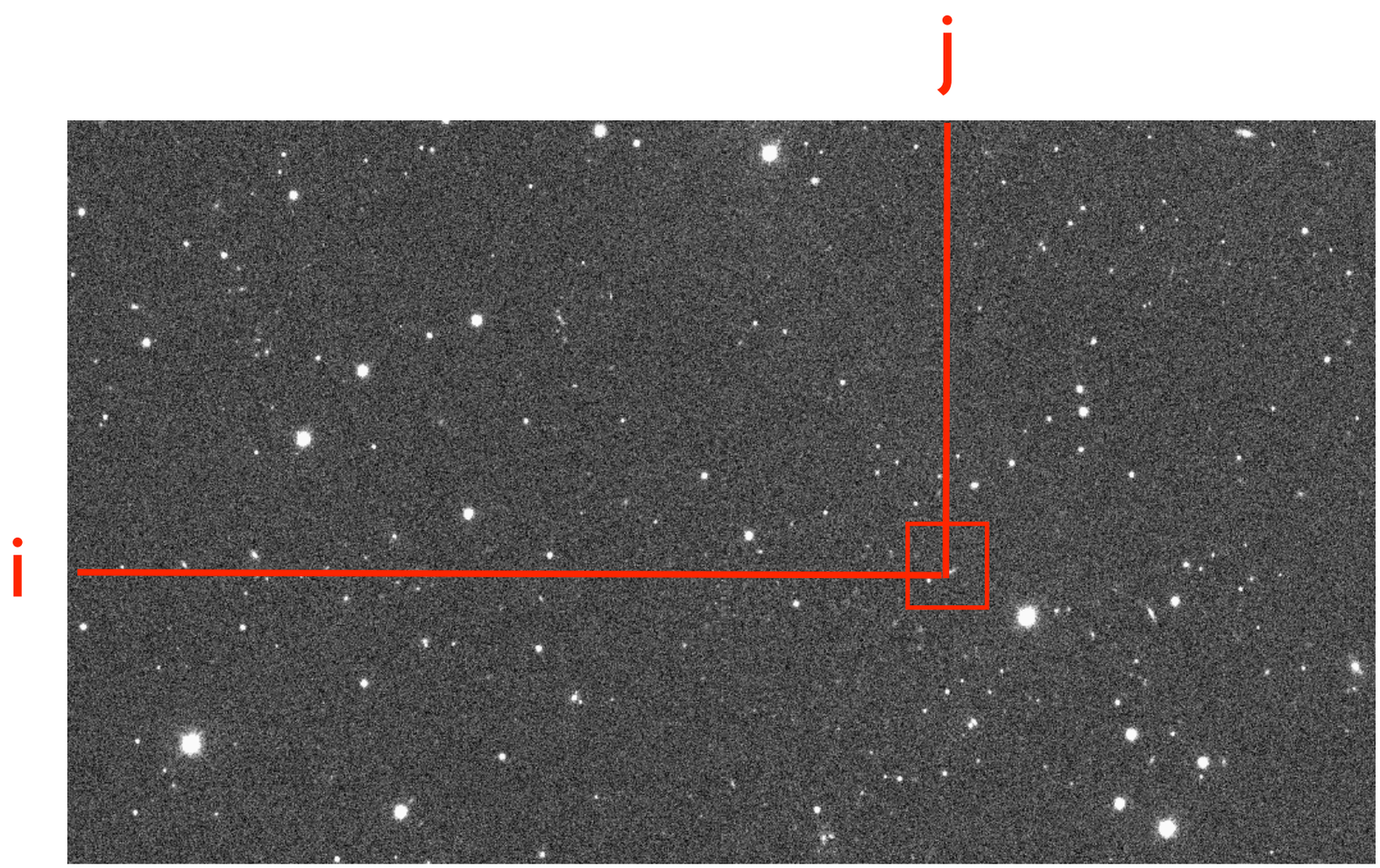
Obtaining Cutouts from SDSS



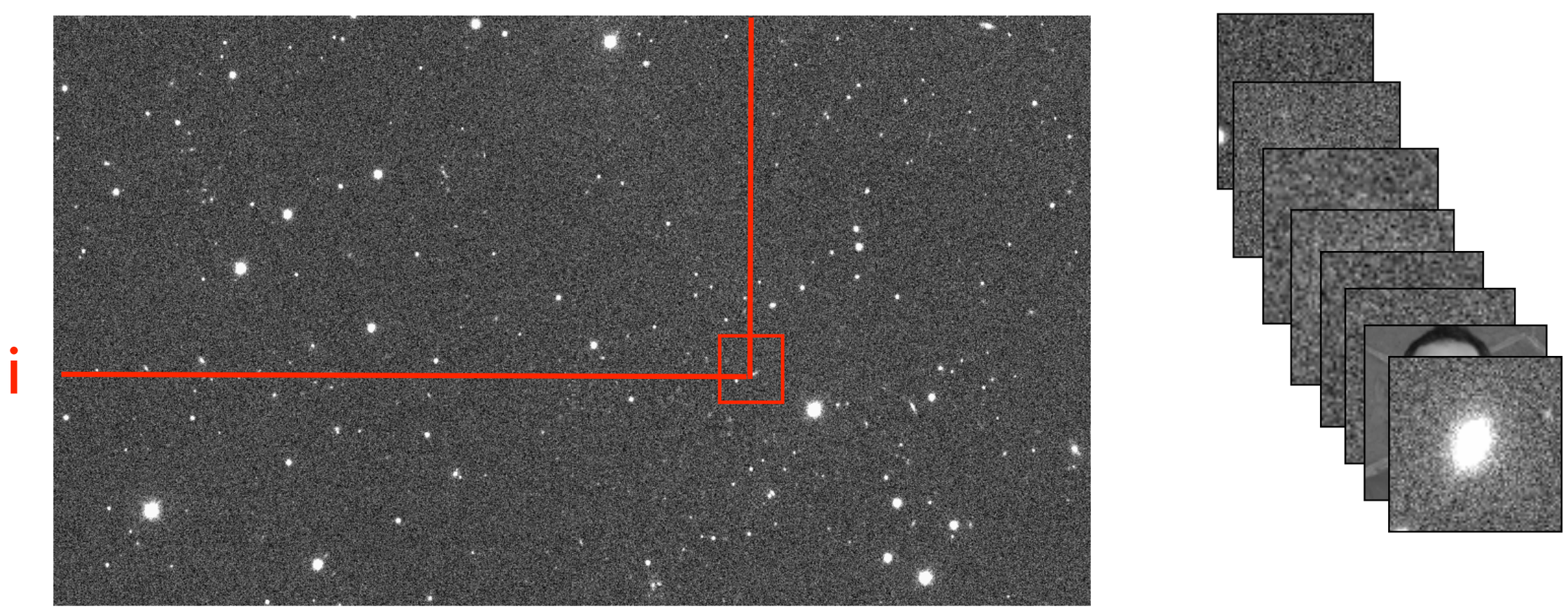
Obtaining Cutouts from SDSS



Obtaining Cutouts from SDSS

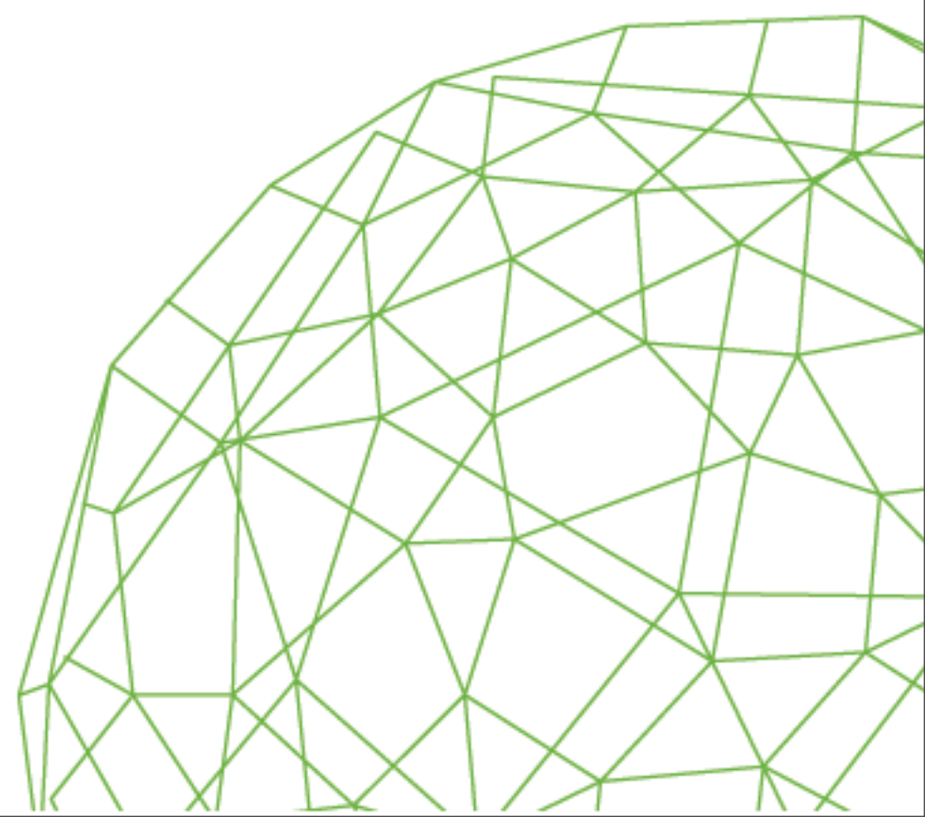


Obtaining Cutouts from SDSS



Procedure

1. Get pixel coordinates from CasJobs
2. Get images from DR7
3. Do a cutout
4. HPC it



Get pixel coordinates from CasJobs

- What do we need?
  - identify the image where the object is by run, rerun, camcol, field.
  - get pixel coordinates
  - get object size: Petrosian Radius
    - Petrosian ratio: ratio of the local surface brightness in an annulus at *r* to the mean surface brightness within *r*.
    - Petrosian radius: radius at which the Petrosian ratio is 0.2.

Get pixel coordinates from CasJobs

- http://casjobs.sdss.org/CasJobs/
- http://casjobs.sdss.org/dr7/en/help/browser/browser.asp

```
select top 100 o.objId, o.rowc_r, o.colc_r, o.petroRad_g,
o.petroRad_r, o.run, o.rerun, o.camcol, o.field, o.petroMag_r
into mydb.obj_row_col
from photoObj as o
where o.petroMag_r >= 16 and o.petroMag_r < 18
```

Get pixel coordinates from CasJobs

TOPCAT(1): Table Browser

Saves one or more tables or the TOPCAT session

Table Browser for 1: obj_row_col_guille.fit

	objId	rowc_r	colc_r	petroRad_g	petroRad_r	run	rerun	camcol	field	petroMag_r
1	588848900971364428	574.64746	1732.50232	1.43801	1.41113	756	44	5	199	16.2626
2	588848900971364430	676.86188	1066.66284	1.43828	1.38989	756	44	5	199	16.26881
3	588848900971364446	951.96545	1969.8999	1.40653	1.39518	756	44	5	199	17.98967
4	588848900971364451	1208.68225	354.98227	1.45333	1.28233	756	44	5	199	16.23476
5	588848900971364457	1270.63232	1098.47229	1.60994	2.19224	756	44	5	199	16.42279
6	588848900971364482	1289.84619	891.37158	1.45655	1.32003	756	44	5	199	17.3641
7	588848900971364484	1327.46729	593.16467	1.48057	1.29462	756	44	5	199	16.34192
8	588848900971364492	149.33058	1614.99695	1.41921	1.45908	756	44	5	199	16.8724
9	588848900971364505	248.26163	1172.94531	1.43005	1.43577	756	44	5	199	17.74933
10	588848900971364509	331.1709	681.07166	1.43374	1.41923	756	44	5	199	17.79399
11	588848900971364520	446.01651	465.50851	1.40243	1.36929	756	44	5	199	17.07066
12	588848900971364522	500.36752	738.21783	1.44192	1.37903	756	44	5	199	17.6296
13	588848900971364524	534.19757	26.55153	1.34153	1.3418	756	44	5	199	17.66015
14	588848900971364534	646.13995	755.28229	1.42393	1.37804	756	44	5	199	17.99287
15	588848900971364538	770.20422	997.11499	1.42436	1.37406	756	44	5	199	17.50831
16	588848900971364549	905.85846	1583.44409	1.47688	1.39276	756	44	5	199	17.31635
17	588848900971364560	1114.78882	246.49637	1.38777	1.29355	756	44	5	199	17.70358
18	588848900971364564	1153.02185	1935.00024	1.44794	1.36904	756	44	5	199	17.77866
19	588848900971364569	1182.45825	23.92278	1.4289	1.266	756	44	5	199	17.09027
20	588848900971364579	1370.04114	1574.55798	1.49369	1.36375	756	44	5	199	16.79921
21	588848900971364580	1378.06689	1573.60364	1.34412	1.2479	756	44	5	199	17.7473
22	588848900971364598	471.71838	2014.63745	8.23456	6.86085	756	44	5	199	17.75352
23	588848900971429915	261.01761	768.92902	1.49208	1.29233	756	44	5	200	16.23797

Get images from DR7

- http://das.sdss.org/imaging/1462/40/corr/4/fpC-001462-r4-0468.fit.gz

RUN RERUN CAMCOL RUN CAMCOL FIELD

```
$ python image_list.py obj_row_col_guille.fit get_files.sh
total =  100
unique =  5
http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0199.fit.gz
http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0200.fit.gz
http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0201.fit.gz
http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0202.fit.gz
http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0203.fit.gz
```

Get images from DR7

- http://das.sdss.org/imaging/1462/40/corr/4/fpC-001462-r4-0468.fit.gz

RUN RERUN CAMCOL RUN CAMCOL FIELD

```
import pyfits as pf
import sys
import numpy as np
from numpy.core.defchararray import add, zfill

tbdata = pf.open(sys.argv[1])[1].data

run    = np.array(map (str, tbdata.field('RUN')))
camcol = np.array(map (str, tbdata.field('CAMCOL')))
rerun  = np.array(map (str, tbdata.field('RERUN')))
field  = np.array(map (str, tbdata.field('FIELD')))

uri = add(add(add(add(add(add(add(add(add(add(add(add("http://das.sdss.org/imaging/",
                                                      run), "/"), rerun), "/corr/"),
                                           camcol), "/fpC-"), zfill(run, 6)), "-r"),
                         camcol), "-"), zfill(field, 4)), ".fit.gz")
print "total = ", len(uri)
uri = np.unique(uri)
print "unique = ", len(uri)

file = open(sys.argv[2], "w")
for i in range(len(uri)):
    print uri[i]
    file.write ("wget -b -c -P " + uri[i].split("/")[5] + " " + uri[i]+"\n")
file.close()
```

get_files.sh

```
wget -b -c -P 44 http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0199.fit.gz
wget -b -c -P 44 http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0200.fit.gz
wget -b -c -P 44 http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0201.fit.gz
wget -b -c -P 44 http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0202.fit.gz
wget -b -c -P 44 http://das.sdss.org/imaging/756/44/corr/5/fpC-000756-r5-0203.fit.gz
```

Get images from DR7

```
$ chmod 755 get_files.sh

$ ./get_files.sh
Continuing in background, pid 63811.
Output will be written to `wget-log'.
Continuing in background, pid 63813.
Output will be written to `wget-log.1'.
Continuing in background, pid 63815.
Output will be written to `wget-log.2'.
Continuing in background, pid 63817.
Output will be written to `wget-log.3'.
Continuing in background, pid 63819.
Output will be written to `wget-log.4'.
```

Do a cutout

```
def cut (hdulist, rc, cc, size):
    size = int(round(size/2) * 2) # integer and multiple of 2
    rc = int(round(rc)) # integer
    cc = int(round(cc)) # integer

    Nr = hdulist[0].data.shape[0]
    Nc = hdulist[0].data.shape[1]
    # borders in original image
    r1 = max([rc - size/2, 0])
    r2 = min([rc + size/2, Nr])
    c1 = max([cc - size/2, 0])
    c2 = min([cc + size/2, Nc])

    # borders in new image considering they may not overlap exactly
    r1e = size/2-(rc-r1)
    r2e = size/2+(r2-rc)
    c1e = size/2-(cc-c1)
    c2e = size/2+(c2-cc)

    data_z = np.zeros((size, size)) # new image
    data_z[r1e:r2e, c1e:c2e] = hdulist[0].data[r1:r2, c1:c2] # set data

    # change header:
    pix1 = hdulist[0].header['CRPIX1']
    pix2 = hdulist[0].header['CRPIX2']
    hdulist[0].header['CRPIX1'] = size/2 - rc + pix1
    hdulist[0].header['CRPIX2'] = size/2 - cc + pix2

    hdulist[0].data = data_z
```

Do a cutout

```
# python ../getCutout.py ../obj_row_col_guille.fit ../ 1
import sys
import os

import pyfits as pf
import numpy as np
from numpy.core.defchararray import add, zfill

tbdata = pf.open (sys.argv[1])[1].data # first cl par: table file
im_path = sys.argv[2]                  # second cl par: images path
i = int(sys.argv[3]) - 1               # third cl par: 1-offset index

# parameters needed
objId  = str (tbdata.field('objId')[i])
run    = str (tbdata.field('RUN')[i])
camcol = str (tbdata.field('CAMCOL')[i])
rerun  = str (tbdata.field('RERUN')[i])
field  = str (tbdata.field('FIELD')[i])
rowc_r = tbdata.field('ROWC_r')[i]
colc_r = tbdata.field('COLC_r')[i]
petroRad_r = tbdata.field('petroRad_r')[i]
filename = (im_path + rerun + "/fpC-" + str(zfill(run, 6)) +
            "-r" + camcol + "-" +  str(zfill(field, 4)) + ".fit.gz")
hdulist = pf.open(filename, output_verify = "silentfix") # open fits file

cut (hdulist, rowc_r, colc_r, 2*petroRad_r/0.4) # 0.4: radius from arcsec to pixels
out_name = objId + "_" + run + "_" + rerun + "_" + camcol + "_" + field + "_r.fits"
hdulist.writeto(out_name, clobber=True, output_verify = "ignore") # write cutout file
```

HPC it: getting ready at scratch

```
$ ssh student01@development.dim.uchile.cl

[student01@development ~]$ mkdir LSSDS_gcabrera

[student01@development ~]$ cd LSSDS_gcabrera

[student01@development LSSDS_gcabrera]$ mkdir cutouts_demo

[student01@development LSSDS_gcabrera]$ cd cutouts_demo/

[student01@development cutouts_demo]$ cp -r ~/HPC_Demo/SDSS_images/ .

[student01@development cutouts_demo]$ cp ~/HPC_Demo/obj_row_col_guille.fit .

[student01@development cutouts_demo]$ cp ~/HPC_Demo/getCutout.py .

[student01@development cutouts_demo]$ mkdir cutouts

[student01@development cutouts_demo]$ cp ~/HPC_Demo/cutouts/sge-cutouts.sh cutouts/

[student01@development cutouts_demo]$ cd cutouts/
```

HPC it: sge script

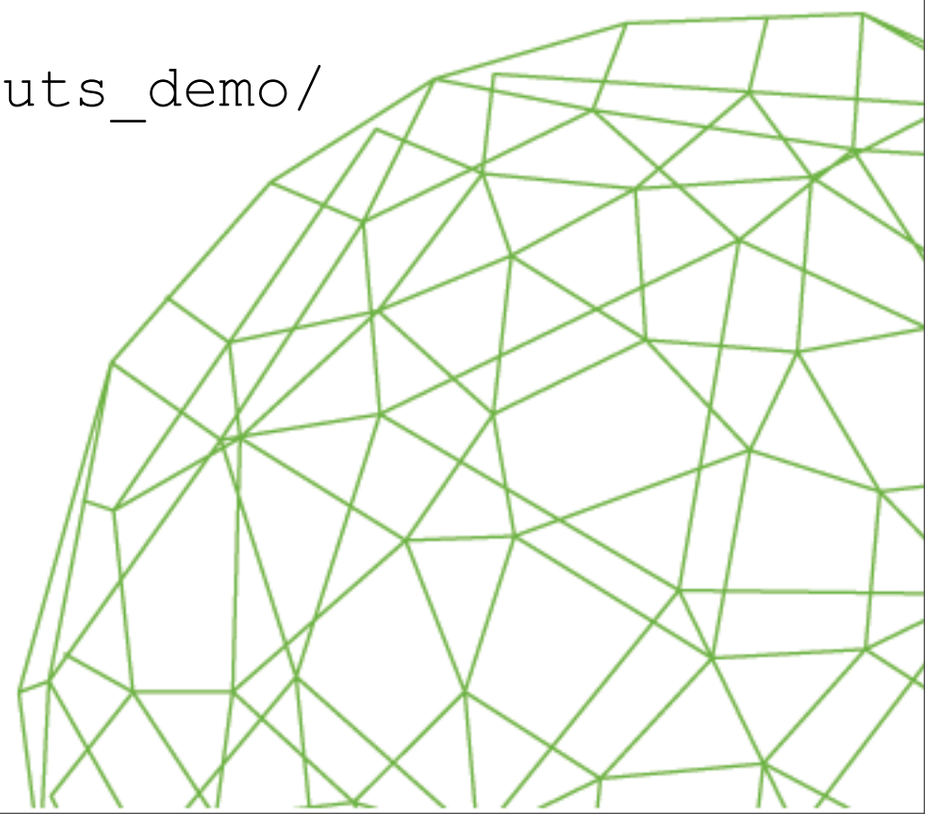
```
#!/bin/bash
#$ -cwd
#$ -j y
#$ -m abes
##$ -M gcabrera@dim.uchile.cl
#$ -N logs_Cutouts
#$ -notify
#$ -S /bin/bash
#$ -q all.q
## Create 100 iterations: $SGE_TASK_ID = 1,...,100
#$ -t 1-100:1

## Variable $JOB_NAME contains the value specified in option -N
## Variable $JOB_ID contains the job ID
## Variable $SGE_TASK_ID contains the sequence number

# Load modules environment(so that command modules work)
. /etc/profile.d/modules.sh

module load astro
#export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/home/local/demo/demo01/.local/lib/

python ../getCutout.py ../obj_row_col_guille.fit ../SDSS_images/ $SGE_TASK_ID
```

HPC it: sge script explanation

```
#$ -cwd -- Place the output files (.e,.o) in the current working directory.
     The default is to place them in the users home directory.

#$ -j y -- Merge the standard out and standard error to one file

#$ -m abes -- Send mail at submission and completion of script
##$ -M gcabrera@dim.uchile.cl

#$ -N cutouts -- The name of the job

#$ -notify -- send a signal to your program some time before the application
is finally killed for some predictable reason.

#$ -S /bin/bash -- Specify the shell to use when running the job script

#$ -q all.q -- The job should be placed into the queue 'all.q'

#$ -t 1-100:1 -- Create 100 iterations: $SGE_TASK_ID = 1,...,100
```

HPC it: run it

```
$ qsub sge-cutouts.sh

$ qstat

$ ls

$ cd ..

$ zip -r cutouts.zip cutouts/

$ cp cutouts.zip ~/LSSDS_gcabrera/cutouts_demo/
```

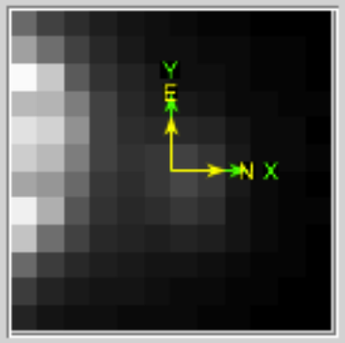


HPC it: get it

```
Guille$ sftp student01@development.cmm.uchile.cl

sftp> cd LSSDS_gcabrera/cutouts_demo/

sftp> get cutouts.zip
Fetching /home/local/demo/demo01/HPC_Demo/cutouts.zip to cutouts.zip
/home/local/demo/demo01/HPC_Demo/cutouts.zip       100%  314KB 314.1KB/s   00:01

sftp> exit

Guille$ unzip cutouts.zip

Guille$ ds9 cutouts/5*
```

SAOImage ds9

Fichero		588848900971626555_756_44_5_203_r.fit		
Objeto		/ e.g.,		
Valor				
WCS				
Física	X		Y	
Imagen	X		Y	
Marco 100	Zoom	1.000	Ángulo	0.000



fichero | editar | ver | marco | bin | zoom | escala | color | región | wcs | ayuda

acerca de | abrir | guardar imagen | samp send | cabecera | formato de página | imprimir | salir