```
!git clone https://github.com/Pabsthegreat/DeepRule-master.git
```

```
Cloning into 'DeepRule-master'...
remote: Enumerating objects: 213, done.
remote: Counting objects: 100% (213/213), done.
remote: Compressing objects: 100% (137/137), done.
remote: Total 213 (delta 65), reused 213 (delta 65), pack-reused 0 (from 0)
Receiving objects: 100% (213/213), 3.81 MiB | 29.10 MiB/s, done.
Resolving deltas: 100% (65/65), done.
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
# 1. Install Miniconda silently into /content/conda
!wget -q https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh -O miniconda.sh
!bash miniconda.sh -b -p /content/conda

# 2. Put conda on PATH for THIS session
import os
os.environ["PATH"] = "/content/conda/bin:" + os.environ["PATH"]

# 3. Create env. Try using their env file if it exists.
#    If this fails, it's fine. We'll do manual install next step.
!conda env create -n DeepRule --file DeepRule.txt || echo "conda env create from DeepRule.txt failed, w

```

```
PREFIX=/content/conda
Unpacking bootstrapper...
Unpacking payload...

Installing base environment...

Preparing transaction: ...working... done
Executing transaction: ...working... done
installation finished.
WARNING:
    You currently have a PYTHONPATH environment variable set. This may cause
    unexpected behavior when running the Python interpreter in Miniconda3.
    For best results, please verify that your PYTHONPATH only points to
    directories of packages that are compatible with the Python interpreter
    in Miniconda3: /content/conda
Jupyter detected...

CondaToSNonInteractiveError: Terms of Service have not been accepted for the following channels. Please acc
    - https://repo.anaconda.com/pkgs/main
    - https://repo.anaconda.com/pkgs/r

To accept these channels' Terms of Service, run the following commands:
    conda tos accept --override-channels --channel https://repo.anaconda.com/pkgs/main
    conda tos accept --override-channels --channel https://repo.anaconda.com/pkgs/r

For information on safely removing channels from your conda configuration,
please see the official documentation:

    https://www.anaconda.com/docs/tools/working-with-conda/channels

conda env create from DeepRule.txt failed, will install manually.
```

```
import os
os.environ["PATH"] = "/content/conda/bin:" + os.environ["PATH"]

!conda create -y -n DeepRule python=3.9

```

```
Jupyter detected...

CondaToSNonInteractiveError: Terms of Service have not been accepted for the following channels. Please acc
    - https://repo.anaconda.com/pkgs/main
    - https://repo.anaconda.com/pkgs/r

To accept these channels' Terms of Service, run the following commands:
    conda tos accept --override-channels --channel https://repo.anaconda.com/pkgs/main
    conda tos accept --override-channels --channel https://repo.anaconda.com/pkgs/r

For information on safely removing channels from your conda configuration,
please see the official documentation:
```

https://www.anaconda.com/docs/tools/working-with-conda/channels

```
!conda tos accept --override-channels --channel https://repo.anaconda.com/pkgs/main
!conda tos accept --override-channels --channel https://repo.anaconda.com/pkgs/r

```

```
accepted Terms of Service for https://repo.anaconda.com/pkgs/main
accepted Terms of Service for https://repo.anaconda.com/pkgs/r
```

```
!conda create -y -n DeepRule python=3.9

```

```
  tzdata             pkgs/main/noarch::tzdata-2025b-h04d1e81_0
  wheel              pkgs/main/linux-64::wheel-0.45.1-py39h06a4308_0
  xorg-libx11        pkgs/main/linux-64::xorg-libx11-1.8.12-h9b100fa_1
  xorg-libxau        pkgs/main/linux-64::xorg-libxau-1.0.12-h9b100fa_0
  xorg-libxdmcp      pkgs/main/linux-64::xorg-libxdmcp-1.1.5-h9b100fa_0
  xorg-xorgproto     pkgs/main/linux-64::xorg-xorgproto-2024.1-h5eee18b_1
  xz                 pkgs/main/linux-64::xz-5.6.4-h5eee18b_1
  zlib               pkgs/main/linux-64::zlib-1.3.1-hb25bd0a_0



Downloading and Extracting Packages:
python-3.9.25        | 23.0 MB   | :   0% 0/1 [00:00<?, ?it/s]
setuptools-80.9.0    | 1.4 MB    | :   0% 0/1 [00:00<?, ?it/s]

wheel-0.45.1         | 114 KB    | :   0% 0/1 [00:00<?, ?it/s]


libnsl-2.0.0         | 31 KB     | :   0% 0/1 [00:00<?, ?it/s]


libnsl-2.0.0         | 31 KB     | :  52% 0.515966492410405/1 [00:00<00:00,  2.86it/s]


libnsl-2.0.0         | 31 KB     | : 100% 1.0/1 [00:00<00:00,  2.86it/s]


libnsl-2.0.0         | 31 KB     | : 100% 1.0/1 [00:00<00:00,  2.86it/s]
setuptools-80.9.0    | 1.4 MB    | :   1% 0.011036459926145175/1 [00:00<00:17, 17.70s/it]

python-3.9.25        | 23.0 MB   | :   0% 0.0006784110506800052/1 [00:00<05:19, 320.20s/it]

wheel-0.45.1         | 114 KB    | : 100% 1.0/1 [00:00<00:00,  1.53s/it]
python-3.9.25        | 23.0 MB   | :  32% 0.3229236601236825/1 [00:00<00:00,  1.28it/s]

wheel-0.45.1         | 114 KB    | : 100% 1.0/1 [00:00<00:00,  3.64it/s]

python-3.9.25        | 23.0 MB   | : 100% 1.0/1 [00:00<00:00,  2.44it/s]
setuptools-80.9.0    | 1.4 MB    | : 100% 1.0/1 [00:00<00:00,  1.48it/s]







Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate DeepRule
#
# To deactivate an active environment, use
#
#     $ conda deactivate
```

```
import os
os.environ["PATH"] = "/content/conda/envs/DeepRule/bin:" + os.environ["PATH"]
os.environ["CONDA_DEFAULT_ENV"] = "DeepRule"

!which python
!python --version

```

```
/content/conda/envs/DeepRule/bin/python
Python 3.9.25
```

```
!conda install -y -n DeepRule pytorch torchvision pytorch-cuda=12.1 -c pytorch -c nvidia
!conda install -y -n DeepRule cython numpy scipy matplotlib scikit-image scikit-learn opencv pip
!conda run -n DeepRule pip install pycocotools tqdm shapely pytesseract

```

```
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
Collecting pycocotools
  Downloading pycocotools-2.0.10-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (1.3 kB
Collecting tqdm
  Downloading tqdm-4.67.1-py3-none-any.whl.metadata (57 kB)
Collecting shapely
  Downloading shapely-2.0.7-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.8 kB)
Collecting pytesseract
  Downloading pytesseract-0.3.13-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: numpy in ./conda/envs/DeepRule/lib/python3.9/site-packages (from pycocotool
Requirement already satisfied: packaging>=21.3 in ./conda/envs/DeepRule/lib/python3.9/site-packages (from
Requirement already satisfied: Pillow>=8.0.0 in ./conda/envs/DeepRule/lib/python3.9/site-packages (from py
Downloading pycocotools-2.0.10-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (453 kB)
Downloading tqdm-4.67.1-py3-none-any.whl (78 kB)
Downloading shapely-2.0.7-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.5 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.5/2.5 MB 81.3 MB/s  0:00:00
Downloading pytesseract-0.3.13-py3-none-any.whl (14 kB)
Installing collected packages: tqdm, shapely, pytesseract, pycocotools

Successfully installed pycocotools-2.0.10 pytesseract-0.3.13 shapely-2.0.7 tqdm-4.67.1
```

```
import os
os.environ["PATH"] = "/content/conda/envs/DeepRule/bin:" + os.environ["PATH"]
os.environ["CONDA_DEFAULT_ENV"] = "DeepRule"

# sanity check
!which python
!python --version
!python -c "import torch, sys; print('torch', torch.__version__, 'cuda?', torch.cuda.is_available(), '|

```

```
/content/conda/envs/DeepRule/bin/python
Python 3.9.25
torch 2.5.1 cuda? True python 3.9.25 (main, Nov  3 2025, 22:33:05)
[GCC 11.2.0]
```

```
!apt-get update
!apt-get install -y tesseract-ocr
!apt-get install -y libtesseract-dev
```

```
4
5 # sanity check
6 !tesseract --version
7
```

```
Hit:11 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Get:12 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [127 kB]
Get:13 https://r2u.stat.illinois.edu/ubuntu jammy/main amd64 Packages [2,822 kB]
Get:14 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 Packages [1,594 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy-updates/restricted amd64 Packages [6,161 kB]
Get:16 http://security.ubuntu.com/ubuntu jammy-security/universe amd64 Packages [1,289 kB]
Get:17 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [3,855 kB]
Get:18 https://r2u.stat.illinois.edu/ubuntu jammy/main all Packages [9,411 kB]
Get:19 http://security.ubuntu.com/ubuntu jammy-security/main amd64 Packages [3,520 kB]
Get:20 http://security.ubuntu.com/ubuntu jammy-security/restricted amd64 Packages [5,963 kB]
Fetched 35.1 MB in 8s (4,607 kB/s)
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as repository 'https://r2u.stat.illinois.edu/
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
tesseract-ocr is already the newest version (4.1.1-2.1build1).
0 upgraded, 0 newly installed, 0 to remove and 47 not upgraded.
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  libarchive-dev libleptonica-dev
The following NEW packages will be installed:
  libarchive-dev libleptonica-dev libtesseract-dev
0 upgraded, 3 newly installed, 0 to remove and 47 not upgraded.
Need to get 3,743 kB of archives.
After this operation, 16.0 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libarchive-dev amd64 3.6.0-1ubuntu1.5 [581
Get:2 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libleptonica-dev amd64 1.82.0-3build1 [1,562 k
Get:3 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libtesseract-dev amd64 4.1.1-2.1build1 [1,600
Fetched 3,743 kB in 2s (1,934 kB/s)
Selecting previously unselected package libarchive-dev:amd64.
(Reading database ... 125080 files and directories currently installed.)
Preparing to unpack .../libarchive-dev_3.6.0-1ubuntu1.5_amd64.deb ...
Unpacking libarchive-dev:amd64 (3.6.0-1ubuntu1.5) ...
Selecting previously unselected package libleptonica-dev.
Preparing to unpack .../libleptonica-dev_1.82.0-3build1_amd64.deb ...
Unpacking libleptonica-dev (1.82.0-3build1) ...
Selecting previously unselected package libtesseract-dev:amd64.
Preparing to unpack .../libtesseract-dev_4.1.1-2.1build1_amd64.deb ...
Unpacking libtesseract-dev:amd64 (4.1.1-2.1build1) ...
Setting up libleptonica-dev (1.82.0-3build1) ...
Setting up libarchive-dev:amd64 (3.6.0-1ubuntu1.5) ...
Setting up libtesseract-dev:amd64 (4.1.1-2.1build1) ...
Processing triggers for man-db (2.10.2-1) ...
tesseract 5.2.0
 leptonica-1.82.0
  libgif 5.2.2 : libjpeg 9f : libpng 1.6.50 : libtiff 4.7.0 : zlib 1.2.13 : libwebp 1.3.2 : libopenjp2 2.5
 Found AVX512BW
 Found AVX512F
 Found AVX2
 Found AVX
 Found FMA
 Found SSE4.1
 Found OpenMP 201511
 Found libarchive 3.8.2 zlib/1.3.1 liblzma/5.6.4 bz2lib/1.0.8 liblz4/1.9.4 libzstd/1.5.7 libxml2/2.13.8 op
```

```
1 %cd /content/DeepRule-master
2 !ls
3
```

```
/content/DeepRule-master
coco_test.py           models                target.png
conda_packagelist.txt  nnet                  templates
config                 ocr.py                test
config.json            OCR_temp.png          test_chart.py
config.py              outputs               testfile
db                     pycocotool            test_pipeline.py
db.sqlite3             README.md             test_pipe_type_cloud.py
DeepRule.txt           requirements-2023.txt test_pipe_type.py
DeepRule.yaml          RuleGroup             test_pipe_type_web.py
external               sample                test.png
LICENSE                server_match          tqdm
logs.txt               static                train_chart.py
manage.py              target_draw.png       utils
```

```
1 # make sure we're still using the DeepRule env inside this cell
2 import os
```

```
import os
os.environ["PATH"] = "/content/conda/envs/DeepRule/bin:" + os.environ["PATH"]
os.environ["CONDA_DEFAULT_ENV"] = "DeepRule"

# try corner pooling build
%cd /content/DeepRule-master/CornerNet/models/py_utils/_cpools
!python setup.py build_ext --inplace

```

```
[Errno 2] No such file or directory: '/content/DeepRule-master/CornerNet/models/py_utils/_cpools'
/content/DeepRule-master
python: can't open file '/content/DeepRule-master/setup.py': [Errno 2] No such file or directory
```

```
%cd /content/DeepRule-master
!ls -R | sed -n '1,200p'

```

```
./models/py_utils/_cpools/src:
bottom_pool.cpp
left_pool.cpp
right_pool.cpp
top_pool.cpp

./nnet:
__init__.py
py_factory.py

./outputs:
read.txt

./pycocotool:
cocoeval.py
coco.py
__init__.py
_mask.c
_mask.cp37-win_amd64.pyd
_mask.cpython-36m-x86_64-linux-gnu.so
_mask.py
mask.py
_mask.pyx

./RuleGroup:
Bar.py
Cls.py
LIneMatch.py
LineOne.py
LineQuiry.py
Pie.py

./sample:
bar.py
chart_pure.py
chart.py
chart_xy.py
cls.py
coco.py
__init__.py
line_cls.py
line.py
pie.py
sku.py
utils.py

./server_match:
__init__.py
settings.py
urls.py
view.py
wsgi.py

./static:
target_draw.png
target.png

./templates:
```

```
import os
os.environ["PATH"] = "/content/conda/envs/DeepRule/bin:" + os.environ["PATH"]
os.environ["CONDA_DEFAULT_ENV"] = "DeepRule"

!which python
!python --version
!python -c "import torch, sys; print('torch', torch.__version__, 'cuda?', torch.cuda.is_available(), '|

```

```
/content/conda/envs/DeepRule/bin/python
Python 3.9.25
torch 2.5.1 cuda? True python 3.9.25 (main, Nov  3 2025, 22:33:05)
[GCC 11.2.0]
```

```
1 import os
2 os.environ["PATH"] = "/content/conda/envs/DeepRule/bin:" + os.environ["PATH"]
3 os.environ["CONDA_DEFAULT_ENV"] = "DeepRule"
4
5 %cd /content/DeepRule-master/models/py_utils/_cpools
6 !python setup.py build_ext --inplace
7
```

```
/content/DeepRule-master/models/py_utils/_cpools
running build_ext
/content/conda/envs/DeepRule/lib/python3.9/site-packages/torch/utils/cpp_extension.py:497: UserWarning: Att
  warnings.warn(msg.format('we could not find ninja.'))
building 'top_pool' extension
creating build/temp.linux-x86_64-cpython-39/src
g++ -pthread -B /content/conda/envs/DeepRule/compiler_compat -Wno-unused-result -Wsign-compare -DNDEBUG -O2
creating build/lib.linux-x86_64-cpython-39
g++ -pthread -B /content/conda/envs/DeepRule/compiler_compat -Wno-unused-result -Wsign-compare -DNDEBUG -O2
building 'bottom_pool' extension
g++ -pthread -B /content/conda/envs/DeepRule/compiler_compat -Wno-unused-result -Wsign-compare -DNDEBUG -O2
g++ -pthread -B /content/conda/envs/DeepRule/compiler_compat -Wno-unused-result -Wsign-compare -DNDEBUG -O2
building 'left_pool' extension
g++ -pthread -B /content/conda/envs/DeepRule/compiler_compat -Wno-unused-result -Wsign-compare -DNDEBUG -O2
g++ -pthread -B /content/conda/envs/DeepRule/compiler_compat -Wno-unused-result -Wsign-compare -DNDEBUG -O2
building 'right_pool' extension
g++ -pthread -B /content/conda/envs/DeepRule/compiler_compat -Wno-unused-result -Wsign-compare -DNDEBUG -O2
g++ -pthread -B /content/conda/envs/DeepRule/compiler_compat -Wno-unused-result -Wsign-compare -DNDEBUG -O2
copying build/lib.linux-x86_64-cpython-39/top_pool.cpython-39-x86_64-linux-gnu.so ->
copying build/lib.linux-x86_64-cpython-39/bottom_pool.cpython-39-x86_64-linux-gnu.so ->
copying build/lib.linux-x86_64-cpython-39/left_pool.cpython-39-x86_64-linux-gnu.so ->
copying build/lib.linux-x86_64-cpython-39/right_pool.cpython-39-x86_64-linux-gnu.so ->
```

```
1 import os
2 os.environ["PATH"] = "/content/conda/envs/DeepRule/bin:" + os.environ["PATH"]
3 os.environ["CONDA_DEFAULT_ENV"] = "DeepRule"
4
5 %cd /content/DeepRule-master/external
6 !make
7
```

```
/content/DeepRule-master/external
python setup.py build_ext --inplace
Compiling nms.pyx because it changed.
[1/1] Cythonizing nms.pyx
nms.c: In function ‘__pyx_pf_3nms_2soft_nms’:
nms.c:6056:33: warning: comparison of integer expressions of different signedness: ‘int’ and ‘unsigned int’
 6056 |       __pyx_t_11 = (__pyx_v_pos < __pyx_v_N);
      |                                 ^
nms.c:6567:33: warning: comparison of integer expressions of different signedness: ‘int’ and ‘unsigned int’
 6567 |       __pyx_t_11 = (__pyx_v_pos < __pyx_v_N);
      |                                 ^
nms.c: In function ‘__pyx_pf_3nms_4soft_nms_merge’:
nms.c:7841:33: warning: comparison of integer expressions of different signedness: ‘int’ and ‘unsigned int’
 7841 |       __pyx_t_11 = (__pyx_v_pos < __pyx_v_N);
      |                                 ^
nms.c:8544:33: warning: comparison of integer expressions of different signedness: ‘int’ and ‘unsigned int’
 8544 |       __pyx_t_11 = (__pyx_v_pos < __pyx_v_N);
      |                                 ^
rm -rf build
```

```
1 %cd /content/DeepRule-master
2 !sed -n '1,200p' ocr.py
3
```

```
            r = requests.post( https://api.ocr.space/parse/image ,
                              files={filename: f},
                              data=payload,
                              )
        return r.content.decode()
    def ocr_result(image_path):
        subscription_key = "ad143190288d40b79483aa0d5c532724"
        vision_base_url = "https://westus2.api.cognitive.microsoft.com/vision/v2.0/"
        ocr_url = vision_base_url + "recognizeText?mode=Printed"
        headers = {'Ocp-Apim-Subscription-Key': subscription_key, 'Content-Type': 'application/octet-stream'}
        params = {'language': 'eng', 'detectOrientation': 'true'}

        image = Image.open(image_path)
        enh_con = ImageEnhance.Contrast(image)
        contrast = 2.0
        image = enh_con.enhance(contrast)
        # image = image.convert('L')
        # image = image.resize((800, 800))
        image.save('OCR_temp.png')
        image_data = open('OCR_temp.png', "rb").read()
        response = requests.post(ocr_url, headers=headers, params=params, data=image_data)
        response.raise_for_status()
        op_location = response.headers['Operation-Location']
        analysis = {}
        while "recognitionResults" not in analysis.keys():
            time.sleep(3)
            binary_content = requests.get(op_location, headers=headers, params=params).content
            analysis = json.loads(binary_content.decode('ascii'))
        line_infos = [region["lines"] for region in analysis["recognitionResults"]]
        word_infos = []
        for line in line_infos:
            for word_metadata in line:
                for word_info in word_metadata["words"]:
                    if 'confidence' in word_info.keys():
                        if word_info['confidence'] == 'Low':
                            continue
                    if word_info['boundingBox'][0] > word_info['boundingBox'][4]:
                        continue
                    word_infos.append(word_info)
        return word_infos
    image_path = 'C:\\work\\evalset_fqa\\vbar\\bitmap\\'
    image_names = os.listdir(image_path)
    for name in ['495.jpg', '151,jpg']:
        image_file_path = os.path.join(image_path, name)
        result = ocr_result(image_file_path)
        print(result)
```

```
%%writefile /content/DeepRule-master/ocr.py
import cv2
import pytesseract

def ocr_result(image_path):
    """
    Local OCR replacement using Tesseract.
    Returns list of dicts with "text" and "boundingBox".
    boundingBox is a flat list of 8 ints [x0,y0, x1,y1, x2,y2, x3,y3].
    """
    img = cv2.imread(image_path)
    data = pytesseract.image_to_data(img, output_type=pytesseract.Output.DICT)

    words = []
    n = len(data["text"])
    for i in range(n):
        raw_text = data["text"][i]
        if raw_text is None:
            continue
        text = raw_text.strip()
        if text == "":
            continue

        x = data["left"][i]
        y = data["top"][i]
        w = data["width"][i]
        h = data["height"][i]

        bbox = [
            x,   y,        # top-left
            x+w, y,        # top-right
            x+w, y+h,      # bottom-right
            x,   y+h       # bottom-left
        ]

        words.append({
```

```
            "text": text,
            "boundingBox": bbox
        })

    return words
```

```
Overwriting /content/DeepRule-master/ocr.py
```

```
# Install pytesseract in the DeepRule environment using conda install
!conda install -y -n DeepRule pytesseract
```

```
Jupyter detected...
2 channel Terms of Service accepted
Channels:
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /content/conda/envs/DeepRule

  added / updated specs:
    - pytesseract


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    pytesseract-0.3.10         |   py39h06a4308_1          38 KB
    ------------------------------------------------------------
                                           Total:          38 KB

The following NEW packages will be INSTALLED:

  pytesseract        pkgs/main/linux-64::pytesseract-0.3.10-py39h06a4308_1



Downloading and Extracting Packages:

Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

```
!mkdir -p /content/samples/bar
!ls /content/samples/bar

```

```
!sed -i '/azureml/d' /content/DeepRule-master/models/CornerNetPureBar.py
!sed -i '/--cache_path/a\    parser.add_argument("--result_path", type=str, default=None)' /content/Dee
!sed -i 's/draw = ImageDraw.Draw(im)/from PIL import Image\n    if isinstance(im, np.ndarray):\n
!sed -i '1i import numpy as np' /content/DeepRule-master/RuleGroup/Bar.py
!sed -i '1i import numpy as np' /content/DeepRule-master/RuleGroup/Bar.py

!conda install -y -n DeepRule h5py
# # or, if conda fails:
!conda run -n DeepRule pip install h5py pandas
```

```
Jupyter detected...
2 channel Terms of Service accepted
Channels:
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /content/conda/envs/DeepRule

  added / updated specs:
    - h5py


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
```

```
    h5py-3.14.0                |   py39he0d80d8_0         1.4 MB
    mpi-1.0                    |             mpich          13 KB
    mpi4py-4.0.3               |   py39hb6b6513_0         951 KB
    mpich-4.1.1                |       hbae89fd_0        20.4 MB
    ------------------------------------------------------------
                                           Total:        22.7 MB

The following NEW packages will be INSTALLED:

  h5py               pkgs/main/linux-64::h5py-3.14.0-py39he0d80d8_0
  mpi                pkgs/main/linux-64::mpi-1.0-mpich
  mpi4py             pkgs/main/linux-64::mpi4py-4.0.3-py39hb6b6513_0
  mpich              pkgs/main/linux-64::mpich-4.1.1-hbae89fd_0



Downloading and Extracting Packages:
mpich-4.1.1          | 20.4 MB   | :   0% 0/1 [00:00<?, ?it/s]
h5py-3.14.0          | 1.4 MB    | :   0% 0/1 [00:00<?, ?it/s]

mpi4py-4.0.3         | 951 KB    | :   0% 0/1 [00:00<?, ?it/s]


mpi-1.0              | 13 KB     | :   0% 0/1 [00:00<?, ?it/s]
h5py-3.14.0          | 1.4 MB    | :   1% 0.011120009773446091/1 [00:00<00:48, 48.98s/it]
h5py-3.14.0          | 1.4 MB    | : 100% 1.0/1 [00:00<00:00, 48.98s/it]
h5py-3.14.0          | 1.4 MB    | : 100% 1.0/1 [00:00<00:00,  1.99it/s]
h5py-3.14.0          | 1.4 MB    | : 100% 1.0/1 [00:00<00:00,  1.99it/s]


mpi-1.0              | 13 KB     | : 100% 1.0/1 [00:00<00:00,  1.46it/s]


mpi-1.0              | 13 KB     | : 100% 1.0/1 [00:00<00:00,  1.46it/s]


mpich-4.1.1          | 20.4 MB   | : 100% 1.0/1 [00:00<00:00,  1.23it/s]
```

```
  1 !mkdir -p /content/samples/bar
  2 !mkdir -p /content/DeepRule-master/bar
  3 !mkdir -p /content/DeepRule-master/bar/annotations
  4 !mkdir -p /content/DeepRule-master/Cls
  5 !mkdir -p /content/DeepRule-master/Cls/annotations
  6 !cp /content/drive/MyDrive/ChartOCR/DataFiles/'instancesBar(1031)_val2019.json' /content/DeepRule-maste
  7 !cp /content/drive/MyDrive/ChartOCR/DataFiles/cls_val2019.pkl /content/DeepRule-master/cache/cls_val20
  8 !cp /content/drive/MyDrive/ChartOCR/DataFiles/'instancesCls(1031)_val2019.json' /content/DeepRule-maste
  9
 10
```

```
cp: cannot create regular file '/content/DeepRule-master/cache/cls_val2019.pkl': No such file or directory
```

```
  1 !conda run -n DeepRule python -c "import pytesseract; import cv2; import numpy as np; import PIL.Image
```

```
OCR: Test123

OK tesseract: 5.2.0
```

```
  1 !apt-get update
  2 !apt-get install -y tesseract-ocr
  3 !conda run -n DeepRule pip install pytesseract
  4
```

```
Hit:1 https://cli.github.com/packages stable InRelease
Hit:2 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
Hit:3 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64  InRelease
Get:4 http://security.ubuntu.com/ubuntu jammy-security InRelease [129 kB]
Hit:5 http://archive.ubuntu.com/ubuntu jammy InRelease
Hit:6 http://archive.ubuntu.com/ubuntu jammy-updates InRelease
Hit:7 http://archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:8 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Hit:9 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy InRelease
Hit:10 https://r2u.stat.illinois.edu/ubuntu jammy InRelease
Hit:11 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Fetched 129 kB in 1s (102 kB/s)
Reading package lists... Done
W: Skipping acquire of configured file 'main/source/Sources' as repository 'https://r2u.stat.illinois.edu/ul
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
tesseract-ocr is already the newest version (4.1.1-2.1build1).
0 upgraded, 0 newly installed, 0 to remove and 47 not upgraded.
```

```
Requirement already satisfied: pytesseract in /content/conda/envs/DeepRule/lib/python3.9/site-packages (0.3
Requirement already satisfied: packaging>=21.3 in /content/conda/envs/DeepRule/lib/python3.9/site-packages
Requirement already satisfied: Pillow>=8.0.0 in /content/conda/envs/DeepRule/lib/python3.9/site-packages (f
```

```
import os
import matplotlib
matplotlib.use("Agg")

os.environ["PATH"] = "/content/conda/envs/DeepRule/bin:" + os.environ["PATH"]
os.environ["CONDA_DEFAULT_ENV"] = "DeepRule"

%cd /content/DeepRule-master


!export CACHE_PATH=/content/DeepRule-master/cache
!mkdir -p /content/output_bar
!cd /content/DeepRule-master && \
CONDA_PREFIX=/content/conda/envs/DeepRule \
TESSDATA_PREFIX=/usr/share/tesseract-ocr/4.00/tessdata \
MPLBACKEND=Agg \
/content/conda/envs/DeepRule/bin/python test_pipe_type_cloud.py \
  --image_path /content/samples/bar \
  --save_path /content/output_bar \
  --type Bar \
  --data_dir /content/DeepRule-master \
  --cache_path /content/DeepRule-master/cache

```

```
/content/DeepRule-master
/content/DeepRule-master/pycocotool/_mask.py:3: UserWarning: pkg_resources is deprecated as an API. See http
  import sys, pkg_resources, imp
<frozen importlib._bootstrap>:228: RuntimeWarning: compiletime version 3.6 of module 'pycocotools._mask' do
<frozen importlib._bootstrap>:228: RuntimeWarning: builtins.type size changed, may indicate binary incompat
loading parameters at iteration: 50000
loading from cache file: /content/DeepRule-master/cache/chart_val2019.pkl
No cache file found...
loading annotations into memory...
/content/DeepRule-master/bar/annotations/instancesBar(1031)_val2019.json
Done (t=0.51s)
creating index...
index created!
6121it [00:00, 9095.96it/s]
loading annotations into memory...
/content/DeepRule-master/bar/annotations/instancesBar(1031)_val2019.json
Done (t=0.56s)
creating index...
index created!
building neural network...
module_file: models.CornerNetPureBar
use kp pure
total parameters: 198592652
loading parameters...
loading model from /content/DeepRule-master/cache/nnet/CornerNetPureBar/CornerNetPureBar_50000.pkl
Traceback (most recent call last):
  File "/content/DeepRule-master/test_pipe_type_cloud.py", line 241, in <module>
    methods = Pre_load_nets(args.type, 0, args.data_dir, args.cache_path)
  File "/content/DeepRule-master/test_pipe_type_cloud.py", line 70, in Pre_load_nets
    db_bar, nnet_bar = load_net(50000, "CornerNetPureBar", data_dir, cache_dir,
  File "/content/DeepRule-master/test_pipe_type_cloud.py", line 61, in load_net
    nnet.load_params(test_iter)
  File "/content/DeepRule-master/nnet/py_factory.py", line 117, in load_params
    with open(cache_file, "rb") as f:
FileNotFoundError: [Errno 2] No such file or directory: '/content/DeepRule-master/cache/nnet/CornerNetPureBa
```

```
!cp /content/drive/MyDrive/ChartOCR/DataFiles/CornerNetPureBar_50000.pkl  /content/DeepRule-master/cach
!mkdir -p /content/DeepRule-master/cache/nnet/CornerNetCls
!cp /content/drive/MyDrive/ChartOCR/DataFiles/CornerNetCls_50000.pkl  /content/DeepRule-master/cache/nn

```

```
%%writefile /content/DeepRule-master/test_pipe_type_cloud.py
#!/usr/bin/env python
import os
import re
import cv2
import json
import math
import torch
```

```
import numpy as np
import argparse
import pandas as pd
from PIL import Image
from tqdm import tqdm

import matplotlib
matplotlib.use("Agg")

from config import system_configs
from nnet.py_factory import NetworkFactory
from db.datasets import datasets
import importlib

from RuleGroup.Cls import GroupCls
from RuleGroup.Bar import GroupBarRaw
from RuleGroup.LineQuiry import GroupQuiryRaw
from RuleGroup.LIneMatch import GroupLineRaw
from RuleGroup.Pie import GroupPie

import pytesseract
torch.backends.cudnn.benchmark = False


# ---------------------------------------------------------------
# Utility helpers
# ---------------------------------------------------------------

def make_dirs(dirs):
    for d in dirs:
        if not os.path.exists(d):
            os.makedirs(d, exist_ok=True)

def _to_py(o):
    if isinstance(o, (np.floating, np.integer)):
        return o.item()
    if isinstance(o, np.ndarray):
        return o.tolist()
    if isinstance(o, dict):
        return {k: _to_py(v) for k, v in o.items()}
    if isinstance(o, (list, tuple)):
        return [_to_py(v) for v in o]
    return o

def _to_box(coords):
    if coords is None:
        return None
    c = np.array(coords).reshape(-1)
    if c.size == 4:
        x1, y1, x2, y2 = map(float, c.tolist())
        return [min(x1, x2), min(y1, y2), max(x1, x2), max(y1, y2)]
    if c.size == 8:
        xs, ys = c[0::2], c[1::2]
        return [float(xs.min()), float(ys.min()), float(xs.max()), float(ys.max())]
    if c.size >= 4:
        x1, y1, x2, y2 = map(float, c[:4].tolist())
        return [min(x1, x2), min(y1, y2), max(x1, x2), max(y1, y2)]
    return None

def boxes_intersection_ratio(box_big, box_small):
    B, S = _to_box(box_big), _to_box(box_small)
    if B is None or S is None:
        return 0.0
    bx1, by1, bx2, by2 = B
    sx1, sy1, sx2, sy2 = S
    inter_x1 = max(bx1, sx1)
    inter_y1 = max(by1, sy1)
    inter_x2 = min(bx2, sx2)
    inter_y2 = min(by2, sy2)
    if inter_x2 <= inter_x1 or inter_y2 <= inter_y1:
        return 0.0
    inter_area = (inter_x2 - inter_x1) * (inter_y2 - inter_y1)
    small_area = max(1.0, (sx2 - sx1) * (sy2 - sy1))
    return inter_area / small_area


# ---------------------------------------------------------------
# Model loading
# ---------------------------------------------------------------
```

```

def load_net(testiter, cfg_name, data_dir, cache_dir, cuda_id=0):
    cfg_file = os.path.join(system_configs.config_dir, cfg_name + ".json")
    with open(cfg_file) as f:
        configs = json.load(f)
    configs["system"].update({
        "snapshot_name": cfg_name,
        "data_dir": data_dir,
        "cache_dir": cache_dir,
        "result_dir": "result_dir",
        "tar_data_dir": "Cls",
    })
    system_configs.update_config(configs["system"])

    split = {"training": system_configs.train_split,
             "validation": system_configs.val_split,
             "testing": system_configs.test_split}["validation"]

    test_iter = system_configs.max_iter if testiter is None else testiter
    print(f"loading parameters at iteration: {test_iter}")
    dataset = system_configs.dataset
    db = datasets[dataset](configs["db"], split)

    nnet = NetworkFactory(db)
    print("loading parameters...")
    nnet.load_params(test_iter)
    if torch.cuda.is_available():
        nnet.cuda(cuda_id)
    nnet.eval_mode()
    return db, nnet

def Pre_load_nets(chart_type, id_cuda, data_dir, cache_dir):
    methods = {}
    db_cls, nnet_cls = load_net(50000, "CornerNetCls", data_dir, cache_dir, id_cuda)
    path_cls = "testfile.test_CornerNetCls"
    testing_cls = importlib.import_module(path_cls).testing
    methods["Cls"] = [db_cls, nnet_cls, testing_cls]

    if chart_type == "Bar":
        db_bar, nnet_bar = load_net(50000, "CornerNetPureBar", data_dir, cache_dir, id_cuda)
        path_bar = "testfile.test_CornerNetPureBar"
        testing_bar = importlib.import_module(path_bar).testing
        methods["Bar"] = [db_bar, nnet_bar, testing_bar]
    return methods


# ------------------------------------------------------------------
# OCR helpers
# ------------------------------------------------------------------

def ocr_result_full_image(image_path):
    os.environ.setdefault("TESSDATA_PREFIX", "/usr/share/tesseract-ocr/4.00/tessdata")
    img = cv2.imread(image_path)
    if img is None:
        raise RuntimeError(f"Cannot read {image_path}")
    rgb = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
    pil = Image.fromarray(rgb)
    data = pytesseract.image_to_data(pil, lang="eng", output_type=pytesseract.Output.DICT)
    words = []
    for i, txt in enumerate(data["text"]):
        txt = (txt or "").strip()
        if not txt:
            continue
        x, y, w, h = int(data["left"][i]), int(data["top"][i]), int(data["width"][i]), int(data["hei
        words.append({"text": txt, "bbox": [x, y, x + w, y + h]})
    return words

def extract_titles_from_clsinfo(cls_info, word_infos):
    out = {}
    for rid in [1, 2, 3]:
        if rid not in cls_info:
            continue
        rbox = _to_box(cls_info[rid])
        captured = [(w["text"], w["bbox"][1], w["bbox"][0]) for w in word_infos if boxes_intersectio
        captured.sort(key=lambda x: (x[1], x[2]))
        if captured:
            out[str(rid)] = " ".join([c[0] for c in captured])
    return out

```

```
def extract_axis_scale_from_clsinfo(image, cls_info):
    if 5 not in cls_info:
        return None, None
    plot = _to_box(cls_info[5])
    if plot is None:
        return None, None
    x1, y1, x2, y2 = map(int, plot)
    strip = image[y1:y2, max(0, x1 - 60):x1]
    if strip.size == 0:
        return None, None
    up = cv2.resize(strip, (0, 0), fx=2, fy=2, interpolation=cv2.INTER_CUBIC)
    data = pytesseract.image_to_data(Image.fromarray(cv2.cvtColor(up, cv2.COLOR_BGR2RGB)),
                                     config="--psm 6", output_type=pytesseract.Output.DICT)
    nums = []
    for t in data["text"]:
        t = re.sub(r"[^0-9.\-]", "", t or "")
        if t and re.search(r"\d", t):
            try: nums.append(float(t))
            except: pass
    if not nums: return None, None
    return min(nums), max(nums)


# ----------------------------------------------------------------------
# Legend / color matching
# ----------------------------------------------------------------------

def rgb_to_hex(rgb):
    return "#{:02x}{:02x}{:02x}".format(*[int(v) for v in rgb])

def bar_color(img, box):
    x1,y1,x2,y2 = [int(v) for v in box]
    roi = img[y1:y2, x1:x2]
    if roi.size == 0: return (0,0,0)
    h,w = roi.shape[:2]
    ys, ye = int(h*0.2), int(h*0.8)
    xs, xe = int(w*0.2), int(w*0.8)
    core = roi[ys:ye, xs:xe]
    bgr = core.reshape(-1,3).mean(axis=0)
    return (bgr[2], bgr[1], bgr[0])

def find_legend_pairs(img, words):
    H, W = img.shape[:2]
    legend = []
    for w in words:
        x1,y1,x2,y2 = w["bbox"]
        if x1 >= int(W*0.65) and len(w["text"]) >= 2:
            cx = max(0, x1 - 12)
            cy = int((y1+y2)/2)
            sx1, sy1 = max(0, cx-6), max(0, cy-6)
            sx2, sy2 = min(W-1, cx+6), min(H-1, cy+6)
            patch = img[sy1:sy2, sx1:sx2]
            if patch.size == 0: continue
            bgr = patch.reshape(-1,3).mean(axis=0)
            rgb = (bgr[2], bgr[1], bgr[0])
            legend.append({"text": w["text"], "rgb": rgb, "hex": rgb_to_hex(rgb)})
    return legend

def closest_legend(rgb, legend):
    if not legend: return None
    bx,by,bz = rgb
    best, best_d = None, 1e9
    for item in legend:
        rx,ry,rz = item["rgb"]
        d = (bx-rx)**2+(by-ry)**2+(bz-rz)**2
        if d<best_d:
            best,best_d=item,d
    return best


# ----------------------------------------------------------------------
# Main per-image pipeline
# ----------------------------------------------------------------------

def run_on_image(image_path, chart_type="Bar"):
    pil_img = Image.open(image_path)
    img = cv2.imread(image_path)
    if img is None:
        raise RuntimeError(f"Could not read {image_path}")
```

```
    with torch.no_grad():
        cls_db, cls_net, cls_fn = methods["Cls"]
        cls_res = cls_fn(img, cls_db, cls_net, debug=False)
        tls, brs = cls_res[1], cls_res[2]
        _, raw_info = GroupCls(pil_img, tls, brs)
        cls_info = {int(k): _to_box(v) for k,v in raw_info.items()}

        words = ocr_result_full_image(image_path)
        titles = extract_titles_from_clsinfo(cls_info, words)
        y_min, y_max = extract_axis_scale_from_clsinfo(img, cls_info)

        bars_raw = []
        if chart_type == "Bar":
            bdb, bnet, bfn = methods["Bar"]
            bres = bfn(img, bdb, bnet, debug=False)
            tls_b, brs_b = bres[0], bres[1]
            bars_raw = GroupBarRaw(img, tls_b, brs_b)

        legend_items = find_legend_pairs(img, words)

        # Fit pixel→value regression
        tick_pairs = []
        if 5 in cls_info:
            x1,y1,x2,y2 = cls_info[5]
            strip = img[int(y1):int(y2), max(0,int(x1)-70):int(x1)]
            data = pytesseract.image_to_data(
                Image.fromarray(cv2.cvtColor(cv2.resize(strip,(0,0),fx=2,fy=2), cv2.COLOR_BGR2RGB)),
                config="--psm 6 -c tessedit_char_whitelist=0123456789.-",
                output_type=pytesseract.Output.DICT)
            for i, txt in enumerate(data["text"]):
                txt = (txt or "").strip()
                if not txt: continue
                try: val=float(re.sub(r"[^0-9.\-]", "", txt))
                except: continue
                oy=int(y1)+int(data["top"][i]/2)+int(data["height"][i]/4)
                tick_pairs.append((oy,val))
        a=b=None
        if len(tick_pairs)>=2:
            ys=np.array([p[0] for p in tick_pairs])
            vs=np.array([p[1] for p in tick_pairs])
            A=np.vstack([ys,np.ones_like(ys)]).T
            a,b=np.linalg.lstsq(A,vs,rcond=None)[0]
        def y_to_val(y): return float(a*y+b) if a is not None else None

        rows=[]
        for i,(x1,y1,x2,y2) in enumerate(bars_raw or []):
            rgb=bar_color(img,(x1,y1,x2,y2))
            match=closest_legend(rgb,legend_items)
            val=y_to_val(y1)
            rows.append({
                "bar_index":i,
                "label":match["text"] if match else "",
                "color":match["hex"] if match else rgb_to_hex(rgb),
                "x1":x1,"y1_top":y1,"x2":x2,"y2_bottom":y2,
                "pixel_height":(y2-y1),
                "value_estimate":None if val is None else round(val,2)
            })
        base=os.path.splitext(os.path.basename(image_path))[0]
        csv_out=os.path.join(args.save_path,f"{base}_table.csv")
        pd.DataFrame(rows).to_csv(csv_out,index=False)
        print("Wrote:",csv_out)

        return {
            "chart_type":chart_type,
            "chart_title_candidates":titles,
            "y_axis_min_est":y_min,
            "y_axis_max_est":y_max,
            "bars_raw":_to_py(bars_raw)
        }


# ---------------------------------------------------------------
# CLI / main
# ---------------------------------------------------------------

def parse_args():
    parser = argparse.ArgumentParser(description="DeepRule ChartOCR (Bar only)")
    parser.add_argument("--image_path",  default="test",
```

```
                        help="Folder containing chart images to process")
    parser.add_argument("--save_path",   default="save",
                        help="Folder where JSON/CSVs will be written")
    parser.add_argument("--type",        default="Bar",
                        help="Chart type hint (currently only 'Bar' supported)")
    parser.add_argument("--data_dir",    default=".",
                        help="Root dir that matches model configs (e.g. /content/DeepRule-master)")
    parser.add_argument("--cache_path",  default="./cache",
                        help="Where pretrained weights/cache live (e.g. /content/DeepRule-master/cac
    parser.add_argument("--result_path", default=None, type=str,
                        help="(Optional) if set, also write a single JSON here")
    return parser.parse_args()


if __name__=="__main__":
    os.environ.setdefault("MPLBACKEND","Agg")
    os.environ.setdefault("TESSDATA_PREFIX","/usr/share/tesseract-ocr/4.00/tessdata")
    args=parse_args()

    methods=Pre_load_nets(args.type,0,args.data_dir,args.cache_path)
    make_dirs([args.save_path])

    all_results={}
    for f in tqdm(sorted(os.listdir(args.image_path))):
        ip=os.path.join(args.image_path,f)
        if not os.path.isfile(ip): continue
        try: res=run_on_image(ip,args.type)
        except Exception as e: res={"error":str(e)}
        all_results[f]=res
        with open(os.path.join(args.save_path,os.path.splitext(f)[0]+"_result.json"),"w") as jf:
            json.dump(_to_py(res),jf,indent=2)
    with open(os.path.join(args.save_path,"all_results.json"),"w") as f:
        json.dump(_to_py(all_results),f,indent=2)
```

```
Overwriting /content/DeepRule-master/test_pipe_type_cloud.py
```

```
import os
import matplotlib
matplotlib.use("Agg")

os.environ["PATH"] = "/content/conda/envs/DeepRule/bin:" + os.environ["PATH"]
os.environ["CONDA_DEFAULT_ENV"] = "DeepRule"

%cd /content/DeepRule-master



!!export CACHE_PATH=/content/DeepRule-master/cache
!mkdir -p /content/output_bar
!cd /content/DeepRule-master && \
CONDA_PREFIX=/content/conda/envs/DeepRule \
TESSDATA_PREFIX=/usr/share/tesseract-ocr/4.00/tessdata \
MPLBACKEND=Agg \
/content/conda/envs/DeepRule/bin/python test_pipe_type_cloud.py \
  --image_path /content/samples/bar \
  --save_path /content/output_bar \
  --type Bar \
  --data_dir /content/DeepRule-master \
  --cache_path /content/DeepRule-master/cache
```

```
/content/DeepRule-master
/content/DeepRule-master/pycocotool/_mask.py:3: UserWarning: pkg_resources is deprecated as an API. See http
  import sys, pkg_resources, imp
<frozen importlib._bootstrap>:228: RuntimeWarning: compiletime version 3.6 of module 'pycocotools._mask' do
<frozen importlib._bootstrap>:228: RuntimeWarning: builtins.type size changed, may indicate binary incompat
loading parameters at iteration: 50000
loading from cache file: /content/DeepRule-master/cache/cls_val2019.pkl
loading annotations into memory...
/content/DeepRule-master/Cls/annotations/instancesCls(1031)_val2019.json
Done (t=0.05s)
creating index...
index created!
loading annotations into memory...
/content/DeepRule-master/Cls/annotations/instancesCls(1031)_val2019.json
Done (t=0.04s)
creating index...
index created!
module_file: models.CornerNetCls
/content/conda/envs/DeepRule/lib/python3.9/site-packages/torch/nn/_reduction.py:51: UserWarning: size_avera
  warnings.warn(warning.format(ret))
```

```
use kp
total parameters: 199649452
loading parameters...
loading model from /content/DeepRule-master/cache/nnet/CornerNetCls/CornerNetCls_50000.pkl
/content/DeepRule-master/nnet/py_factory.py:119: FutureWarning: You are using `torch.load` with `weights_on
  params = torch.load(f)
loading parameters at iteration: 50000
loading from cache file: /content/DeepRule-master/cache/chart_val2019.pkl
loading annotations into memory...
/content/DeepRule-master/bar/annotations/instancesBar(1031)_val2019.json
Done (t=0.89s)
creating index...
index created!
module_file: models.CornerNetPureBar
use kp pure
total parameters: 198592652
loading parameters...
loading model from /content/DeepRule-master/cache/nnet/CornerNetPureBar/CornerNetPureBar_50000.pkl
  0% 0/1 [00:00<?, ?it/s]tensor([[8.8536, 0.0000, 0.1777]], device='cuda:0')
Wrote: /content/output_bar/8a2b5419565a27b03f57c4ab22f2c78a_d3d3LnByb2plY3RzYnlqZW4uY29tCTE4MS4yMjQuMTM2LjQ
100% 1/1 [00:01<00:00,  1.79s/it]
```

```python
import json
import math
import pandas as pd

# path to the per-image result json you showed me
in_json = "/content/output_bar/8a2da1d1dcbc96b480bc98097f85309b_d3d3Lmdvdi51awkxNTEuMTAxLjUyLjE0NA==.

with open(in_json, "r") as f:
    result = json.load(f)

bars = result["bars_raw"]
y_min_val = result["y_axis_min_est"]
y_max_val = result["y_axis_max_est"]

# If y_min == y_max, scaling failed; we'll mark values as None later
same_scale = (abs(y_max_val - y_min_val) < 1e-6)

# 1. turn bars into rows with geometry
rows = []
for idx, (x1, y1, x2, y2) in enumerate(bars):
    x_center = 0.5 * (x1 + x2)
    pixel_height = (y2 - y1)
    rows.append({
        "bar_index": idx,
        "x1": x1,
        "y1_top": y1,
        "x2": x2,
        "y2_bottom": y2,
        "pixel_height": pixel_height,
        "x_center": x_center
    })

df = pd.DataFrame(rows)

# 2. sort by visual order (left -> right)
df = df.sort_values("x_center").reset_index(drop=True)

# 3. estimate numeric data value for each bar
# We assume taller bar == closer to y_min pixel, etc.
# Need the pixel range for plotting region. Approximate from bars:
plot_top_px    = df["y1_top"].min()
plot_bottom_px = df["y2_bottom"].max()
pixel_range    = plot_bottom_px - plot_top_px  # how tall the chart area is in pixels
value_range    = y_max_val - y_min_val         # data range

def pixel_to_value(y_top):
    """
    Map the top of a bar to a numeric value by linear interpolation.
    Higher bar (smaller y_top) -> larger value.
    """
    if same_scale or pixel_range == 0:
        return None
    # normalized_height = (plot_bottom_px - y_top) / pixel_range in [0..1]
    norm_h = (plot_bottom_px - y_top) / pixel_range
    return y_min_val + norm_h * value_range

df["value_estimate"] = df["y1_top"].apply(pixel_to_value)

```

```
print("Parsed table:")
print(df[["bar_index","pixel_height","value_estimate"]])

# 4. save CSV
out_csv = in_json.replace("_result.json", "_table.csv")
df.to_csv(out_csv, index=False)
print("Wrote:", out_csv)

```

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
/tmp/ipython-input-103681994.py in <cell line: 0>()
      6 in_json =
"/content/output_bar/8a2da1d1dcbc96b480bc98097f85309b_d3d3Lmdvdi51awkxNTEuMTAxLjUyLjE0NA==.xls-2-
0_result.json"
      7
----> 8 with open(in_json, "r") as f:
      9     result = json.load(f)
     10

FileNotFoundError: [Errno 2] No such file or directory:
'/content/output_bar/8a2da1d1dcbc96b480bc98097f85309b_d3d3Lmdvdi51awkxNTEuMTAxLjUyLjE0NA==.xls-2-
0_result.json'
```

Next steps: Explain error