

```
In [12]: import json
from pathlib import Path

import matplotlib.pyplot as plt
import pandas as pd

# Set options to display all rows and columns
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
```

Task 1 Results

```
In [14]: df = pd.read_csv("../results/task1/task1_results_summary.csv") \
    .sort_values(by=["map", "dataset"], ascending=[False, True])
from IPython.display import HTML
HTML(df.to_html(index=False))
```

Out[14]: dataset	prompt	box_threshold	text_threshold	map
C	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.5	0.169622
A	The observation is considered malignant if it shows a cluster of fine, pleomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.	0.3	0.5	0.168362
A	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse, scattered, bilateral calcifications are a benign finding.	0.3	0.3	0.166108
C	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.	0.3	0.3	0.163700
C	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.	0.3	0.3	0.162244
B	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.	0.3	0.7	0.161460
B	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.	0.3	0.7	0.157862
B	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.3	0.157386
A	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.3	0.155958
A	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.	0.3	0.3	0.155412

dataset	prompt	box_threshold	text_threshold	map
C	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.	0.3	0.3	0.154642
C	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse, scattered, bilateral calcifications are a benign finding.	0.3	0.7	0.153564
B	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.	0.3	0.3	0.153494
B	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse, scattered, bilateral calcifications are a benign finding.	0.3	0.3	0.153158
C	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.	0.3	0.5	0.151492
B	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.	0.3	0.3	0.151240
C	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.	0.3	0.7	0.149980
A	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.	0.3	0.3	0.148986
A	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.7	0.145836
A	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.	0.3	0.5	0.145374

dataset	prompt	box_threshold	text_threshold	map
A	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.	0.3	0.5	0.145080
B	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.	0.3	0.5	0.142406
C	The observation is considered malignant if it shows a cluster of fine, pleomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.	0.3	0.7	0.142000
B	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.5	0.133166
C	The observation is considered malignant if it shows a cluster of fine, pleomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.	0.3	0.5	0.130646
C	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.3	0.127048
A	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.3	0.126516
A	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.7	0.126334
B	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse, scattered, bilateral calcifications are a benign finding.	0.3	0.5	0.123618
A	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.5	0.118522

dataset	prompt	box_threshold	text_threshold	map
A	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse, scattered, bilateral calcifications are a benign finding.	0.3	0.7	0.115554
A	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.7	0.115022
A	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse, scattered, bilateral calcifications are a benign finding.	0.3	0.5	0.110864
C	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.	0.3	0.5	0.110360
C	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.7	0.110304
C	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.	0.3	0.7	0.109548
C	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.	0.3	0.7	0.107448
C	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.7	0.105362
B	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.5	0.101288
B	The observation is considered malignant if it shows a cluster of fine, pleiomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.	0.3	0.7	0.100490
C	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse,	0.3	0.5	0.097172

dataset	prompt	box_threshold	text_threshold	map
	scattered, bilateral calcifications are a benign finding.			
B	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.	0.3	0.7	0.097088
B	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse, scattered, bilateral calcifications are a benign finding.	0.3	0.7	0.087806
B	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.	0.3	0.7	0.086966
C	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.3	0.086658
A	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.3	0.084068
B	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.5	0.083144
C	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.5	0.081660
C	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.	0.3	0.5	0.081590
B	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.7	0.080820
A	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.5	0.080596

dataset		prompt	box_threshold	text_threshold	map
A	The observation is considered malignant if it shows a cluster of fine, pleomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.		0.3	0.7	0.077964
B	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.		0.3	0.3	0.076550
C	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.		0.3	0.7	0.076396
C	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.		0.3	0.3	0.076200
B	The observation is considered malignant if it shows a cluster of fine, pleomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.		0.3	0.3	0.074534
C	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.		0.3	0.7	0.074506
A	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.		0.3	0.7	0.074296
A	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.		0.3	0.7	0.074268
B	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.		0.3	0.5	0.070656
A	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.		0.3	0.7	0.069060

dataset	prompt	box_threshold	text_threshold	map
C	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.5	0.068122
B	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.	0.3	0.3	0.068052
B	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.	0.3	0.5	0.068024
C	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.	0.3	0.5	0.067324
A	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.	0.3	0.5	0.059624
B	The observation is considered malignant if it shows a cluster of fine, pleomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.	0.3	0.5	0.058728
B	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.	0.3	0.5	0.057594
C	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.	0.3	0.3	0.055606
A	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.	0.3	0.3	0.054598
C	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.	0.3	0.5	0.054206

dataset	prompt	box_threshold	text_threshold	map
A	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.	0.3	0.7	0.052428
A	Architectural distortion or skin retraction is a condition for a malignant tumor, while a mass that is stable and unchanged for several years is a benign condition.	0.3	0.7	0.048452
C	A segmental or linear-branching distribution of calcifications indicates malignancy, whereas diffuse, scattered, bilateral calcifications are a benign finding.	0.3	0.3	0.047318
B	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.	0.3	0.7	0.046534
C	The observation is considered malignant if it shows a cluster of fine, pleomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.	0.3	0.3	0.043944
B	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.3	0.039618
B	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.7	0.039240
B	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.3	0.039058
A	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.	0.3	0.5	0.038428
C	The condition is benign if ultrasound confirms a simple cyst, but the condition is suspicious for malignancy if the mass is solid with indistinct margins.	0.3	0.3	0.037406

dataset	prompt	box_threshold	text_threshold	map
A	A radiolucent (dark) mass, such as a lipoma or oil cyst, is a benign condition, while a hyperdense (white) and spiculated mass is a malignant condition.	0.3	0.3	0.036748
A	A well-defined or circumscribed mass is observed in benign cases; an indistinct, microlobulated, or obscured mass is observed in malignant cases.	0.3	0.3	0.035348
C	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.7	0.033766
B	The presence of benign vascular calcifications (parallel 'tram-tracks') is not suspicious; the presence of clustered, amorphous calcifications is suspicious for malignancy.	0.3	0.3	0.033276
A	The observation is considered malignant if it shows a cluster of fine, pleomorphic microcalcifications, but it is benign if the calcifications are coarse, large, and popcorn-like.	0.3	0.3	0.032436
B	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.	0.3	0.5	0.032394
A	A mass with smooth, circumscribed margins is a condition for a benign finding, whereas a mass with spiculated, irregular margins is a condition for malignancy.	0.3	0.5	0.031554
A	An observation is benign if it's a round, fat-containing, low-density mass; conversely, it is malignant if it's an irregular, high-density mass.	0.3	0.5	0.031050
B	A small, oval mass with a lucent center (fatty hilum) is a benign intramammary lymph node; a new, developing focal asymmetry is a malignant sign.	0.3	0.7	0.030588

Task 2 - 1: COOP

Training

```
In [19]: train_dir = Path("../results/task2-1/train")
training_curves = {}

for json_path in sorted(train_dir.glob("dataset_*.json")):
    dataset_id = json_path.stem.split("_")[-1]
    with open(json_path, "r") as f:
        payload = json.load(f)
    curve_df = pd.DataFrame(payload.get("data_points", []))
    training_curves[dataset_id] = {
        "model": payload.get("model", "Unknown"),
        "metric": payload.get("metric", "Loss"),
        "data": curve_df,
    }

# training_curves
```

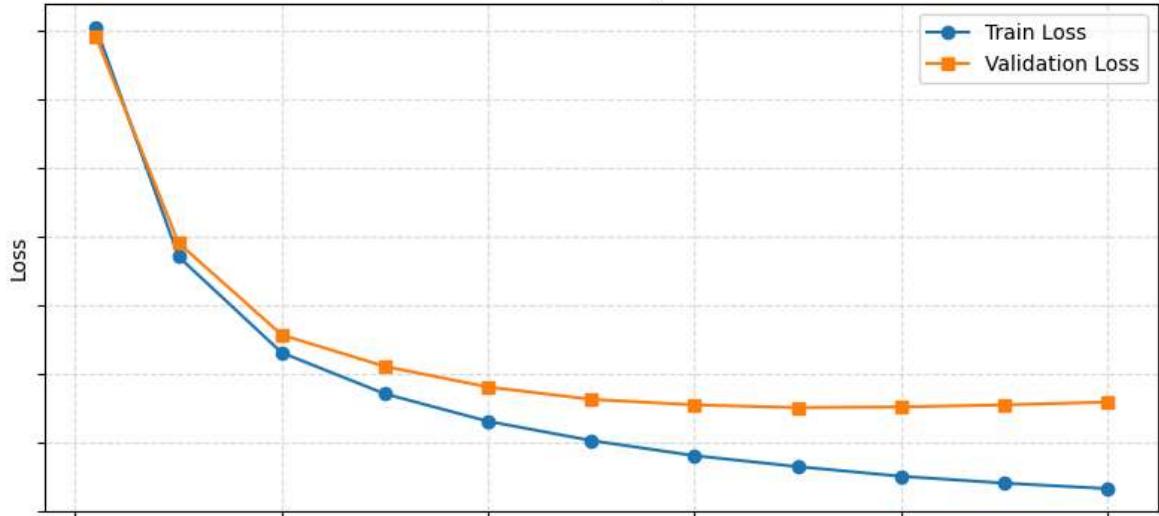
```
In [20]: num_plots = len(training_curves)
fig, axes = plt.subplots(num_plots, 1, figsize=(8, 4 * num_plots), sharex=True)

if num_plots == 1:
    axes = [axes]

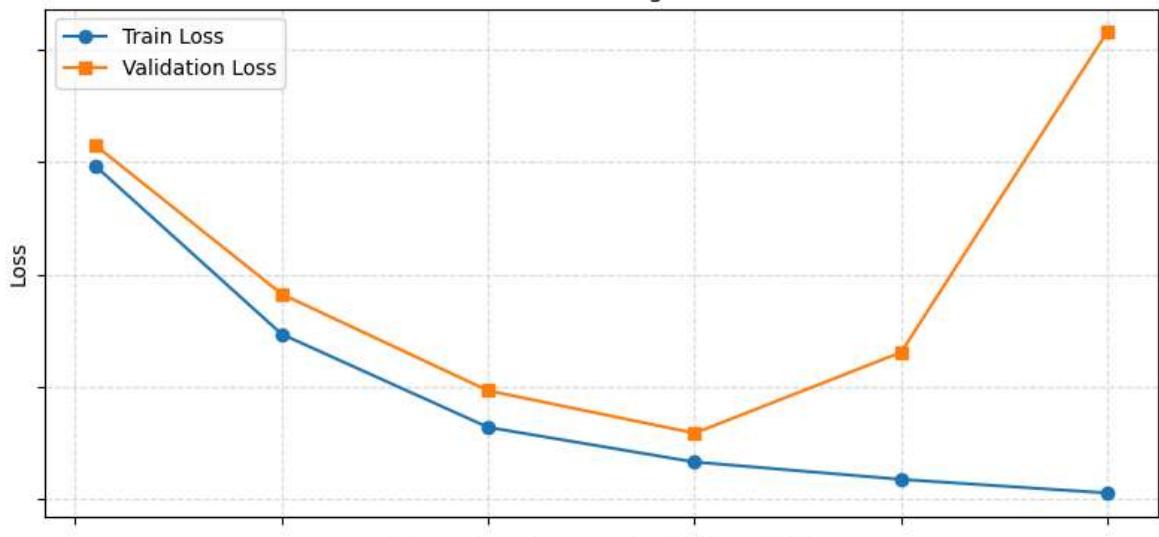
for ax, (dataset_id, payload) in zip(axes, sorted(training_curves.items())):
    curve_df = payload["data"]
    ax.plot(curve_df["epoch"], curve_df["train_loss"], label="Train Loss", marker="o")
    ax.plot(curve_df["epoch"], curve_df["validation_loss"], label="Validation Loss", marker="x")
    ax.set_title(f"Dataset {dataset_id} - {payload['model']}")
    ax.set_ylabel(payload["metric"])
    ax.grid(True, linestyle="--", alpha=0.4)
    ax.legend()
    ax.tick_params(axis="y", labelleft=False)

axes[-1].set_xlabel("Epoch")
plt.tight_layout()
plt.show()
```

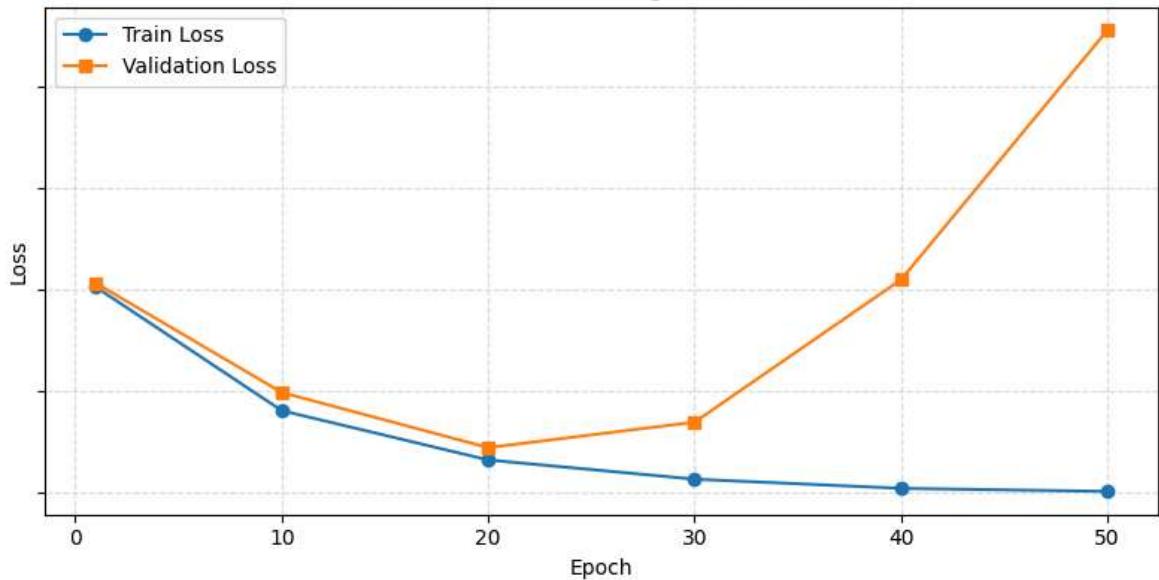
Dataset A - Grounding DINO COOP



Dataset B - Grounding DINO COOP



Dataset C - Grounding DINO COOP



Results

```
In [17]: df = pd.read_csv("../results/task2-1/test/task2_1_results_summary.csv") \
    .sort_values(by=["map", "dataset"], ascending=[False, True])
```

```
from IPython.display import HTML
HTML(df.to_html(index=False))
```

Out[17]:

dataset	box_threshold	text_threshold	model_name	map
B	0.7	0.5	coop_model_B.pth	0.193896
C	0.5	0.5	coop_model_C.pth	0.185178
B	0.3	0.9	coop_model_B.pth	0.183872
C	0.5	0.3	coop_model_C.pth	0.176133
B	0.3	0.3	coop_model_B.pth	0.176014
C	0.5	0.9	coop_model_C.pth	0.174731
C	0.5	0.5	coop_model_B.pth	0.172112
B	0.7	0.5	coop_model_C.pth	0.171003
B	0.7	0.5	coop_model_A.pth	0.169769
C	0.5	0.5	coop_model_A.pth	0.168773
B	0.3	0.9	coop_model_A.pth	0.166901
C	0.5	0.9	coop_model_A.pth	0.166579
B	0.3	0.9	coop_model_C.pth	0.166263
C	0.5	0.9	coop_model_B.pth	0.165955
C	0.5	0.3	coop_model_A.pth	0.161662
C	0.3	0.7	coop_model_C.pth	0.161249
C	0.5	0.3	coop_model_B.pth	0.158250
C	0.5	0.7	coop_model_C.pth	0.157777
C	0.7	0.7	coop_model_C.pth	0.156940
A	0.3	0.9	coop_model_A.pth	0.155201
B	0.3	0.3	coop_model_A.pth	0.151566
B	0.3	0.3	coop_model_C.pth	0.151381
C	0.5	0.7	coop_model_A.pth	0.150993
B	0.7	0.9	coop_model_B.pth	0.149906
C	0.5	0.7	coop_model_B.pth	0.146095
B	0.7	0.9	coop_model_A.pth	0.143435
A	0.3	0.9	coop_model_B.pth	0.142034
C	0.3	0.7	coop_model_B.pth	0.141776
B	0.7	0.9	coop_model_C.pth	0.141509
C	0.3	0.7	coop_model_A.pth	0.141332
A	0.3	0.9	coop_model_C.pth	0.140379
A	0.7	0.7	coop_model_A.pth	0.140351
C	0.7	0.7	coop_model_A.pth	0.139311

dataset	box_threshold	text_threshold	model_name	map
C	0.7	0.7	coop_model_B.pth	0.137321
A	0.5	0.3	coop_model_A.pth	0.135600
A	0.7	0.7	coop_model_B.pth	0.134381
A	0.7	0.7	coop_model_C.pth	0.133110
A	0.3	0.5	coop_model_A.pth	0.131604
C	0.9	0.7	coop_model_C.pth	0.128534
C	0.7	0.3	coop_model_C.pth	0.128442
A	0.3	0.5	coop_model_C.pth	0.124140
C	0.7	0.3	coop_model_A.pth	0.122203
C	0.9	0.7	coop_model_B.pth	0.121352
C	0.7	0.3	coop_model_B.pth	0.121220
C	0.9	0.7	coop_model_A.pth	0.120754
A	0.3	0.5	coop_model_B.pth	0.120611
A	0.5	0.3	coop_model_B.pth	0.120205
A	0.5	0.3	coop_model_C.pth	0.118768
C	0.9	0.5	coop_model_C.pth	0.114890
B	0.5	0.9	coop_model_B.pth	0.114634
C	0.3	0.5	coop_model_C.pth	0.112575
A	0.9	0.7	coop_model_A.pth	0.108390
B	0.3	0.5	coop_model_B.pth	0.106046
A	0.9	0.3	coop_model_A.pth	0.104679
C	0.9	0.5	coop_model_A.pth	0.103459
A	0.9	0.5	coop_model_A.pth	0.103024
C	0.3	0.5	coop_model_A.pth	0.102263
A	0.9	0.7	coop_model_B.pth	0.101821
C	0.3	0.5	coop_model_B.pth	0.101790
A	0.9	0.9	coop_model_A.pth	0.100677
C	0.9	0.5	coop_model_B.pth	0.100187
A	0.9	0.5	coop_model_B.pth	0.099096
A	0.9	0.5	coop_model_C.pth	0.099086
B	0.5	0.9	coop_model_A.pth	0.099067
B	0.7	0.7	coop_model_B.pth	0.098563
B	0.5	0.9	coop_model_C.pth	0.098267

dataset	box_threshold	text_threshold	model_name	map
A	0.9	0.7	coop_model_C.pth	0.098003
B	0.3	0.5	coop_model_C.pth	0.095570
B	0.3	0.5	coop_model_A.pth	0.094863
A	0.9	0.3	coop_model_C.pth	0.093123
A	0.9	0.3	coop_model_B.pth	0.092729
A	0.7	0.3	coop_model_A.pth	0.092421
B	0.5	0.7	coop_model_B.pth	0.091476
A	0.5	0.5	coop_model_A.pth	0.091416
B	0.9	0.3	coop_model_B.pth	0.089635
C	0.9	0.9	coop_model_C.pth	0.089589
A	0.9	0.9	coop_model_B.pth	0.087874
B	0.7	0.7	coop_model_A.pth	0.087281
A	0.9	0.9	coop_model_C.pth	0.086675
A	0.5	0.5	coop_model_C.pth	0.086663
A	0.7	0.3	coop_model_B.pth	0.086463
B	0.9	0.3	coop_model_C.pth	0.086334
B	0.7	0.7	coop_model_C.pth	0.086256
C	0.3	0.3	coop_model_C.pth	0.086080
B	0.9	0.3	coop_model_A.pth	0.085316
C	0.9	0.9	coop_model_B.pth	0.085280
A	0.5	0.5	coop_model_B.pth	0.085210
A	0.7	0.3	coop_model_C.pth	0.084667
C	0.9	0.9	coop_model_A.pth	0.083417
A	0.3	0.3	coop_model_A.pth	0.081219
B	0.5	0.7	coop_model_C.pth	0.081011
B	0.5	0.7	coop_model_A.pth	0.079415
C	0.3	0.3	coop_model_A.pth	0.077985
A	0.7	0.5	coop_model_A.pth	0.077971
C	0.3	0.3	coop_model_B.pth	0.077815
A	0.3	0.3	coop_model_C.pth	0.076769
A	0.3	0.3	coop_model_B.pth	0.075368
A	0.5	0.9	coop_model_A.pth	0.075160
B	0.9	0.5	coop_model_B.pth	0.074925

dataset	box_threshold	text_threshold	model_name	map
A	0.7	0.5	coop_model_C.pth	0.073710
A	0.7	0.5	coop_model_B.pth	0.072714
B	0.9	0.5	coop_model_C.pth	0.070783
B	0.9	0.5	coop_model_A.pth	0.068940
A	0.5	0.9	coop_model_C.pth	0.068510
A	0.5	0.9	coop_model_B.pth	0.068100
A	0.7	0.9	coop_model_A.pth	0.067768
B	0.5	0.3	coop_model_B.pth	0.061595
A	0.7	0.9	coop_model_B.pth	0.061226
C	0.7	0.9	coop_model_C.pth	0.059648
A	0.7	0.9	coop_model_C.pth	0.059599
A	0.5	0.7	coop_model_A.pth	0.059417
C	0.7	0.5	coop_model_C.pth	0.056864
B	0.5	0.3	coop_model_C.pth	0.055443
B	0.5	0.3	coop_model_A.pth	0.055208
C	0.7	0.9	coop_model_A.pth	0.054441
C	0.7	0.9	coop_model_B.pth	0.054013
A	0.5	0.7	coop_model_B.pth	0.052703
A	0.3	0.7	coop_model_A.pth	0.051075
A	0.5	0.7	coop_model_C.pth	0.051026
C	0.7	0.5	coop_model_B.pth	0.049813
C	0.7	0.5	coop_model_A.pth	0.049585
B	0.5	0.5	coop_model_B.pth	0.046615
B	0.9	0.9	coop_model_B.pth	0.045324
C	0.9	0.3	coop_model_C.pth	0.045202
A	0.3	0.7	coop_model_C.pth	0.044970
A	0.3	0.7	coop_model_B.pth	0.044023
B	0.5	0.5	coop_model_C.pth	0.043645
B	0.5	0.5	coop_model_A.pth	0.043070
B	0.9	0.7	coop_model_B.pth	0.041167
C	0.9	0.3	coop_model_B.pth	0.040767
C	0.9	0.3	coop_model_A.pth	0.039854
B	0.9	0.9	coop_model_A.pth	0.039765

dataset	box_threshold	text_threshold	model_name	map
B	0.7	0.3	coop_model_B.pth	0.039764
C	0.3	0.9	coop_model_C.pth	0.039740
B	0.9	0.9	coop_model_C.pth	0.039488
C	0.3	0.9	coop_model_B.pth	0.035949
B	0.9	0.7	coop_model_C.pth	0.035828
B	0.9	0.7	coop_model_A.pth	0.035445
B	0.7	0.3	coop_model_A.pth	0.035085
B	0.7	0.3	coop_model_C.pth	0.034915
B	0.3	0.7	coop_model_B.pth	0.034829
C	0.3	0.9	coop_model_A.pth	0.034733
B	0.3	0.7	coop_model_A.pth	0.032154
B	0.3	0.7	coop_model_C.pth	0.031844

Task 2 - 1: CoCOOP

Training

```
In [23]: cocoop_train_dir = Path("../results/task2-2/train")
cocoop_training_curves = {}

for json_path in sorted(cocoop_train_dir.glob("dataset_*.json")):
    dataset_id = json_path.stem.split("_")[-1].upper()
    with open(json_path, "r", encoding="utf-8") as f:
        records = json.load(f)
    curve_df = pd.DataFrame(records)
    cocoop_training_curves[dataset_id] = curve_df

cocoop_training_curves.keys()
```

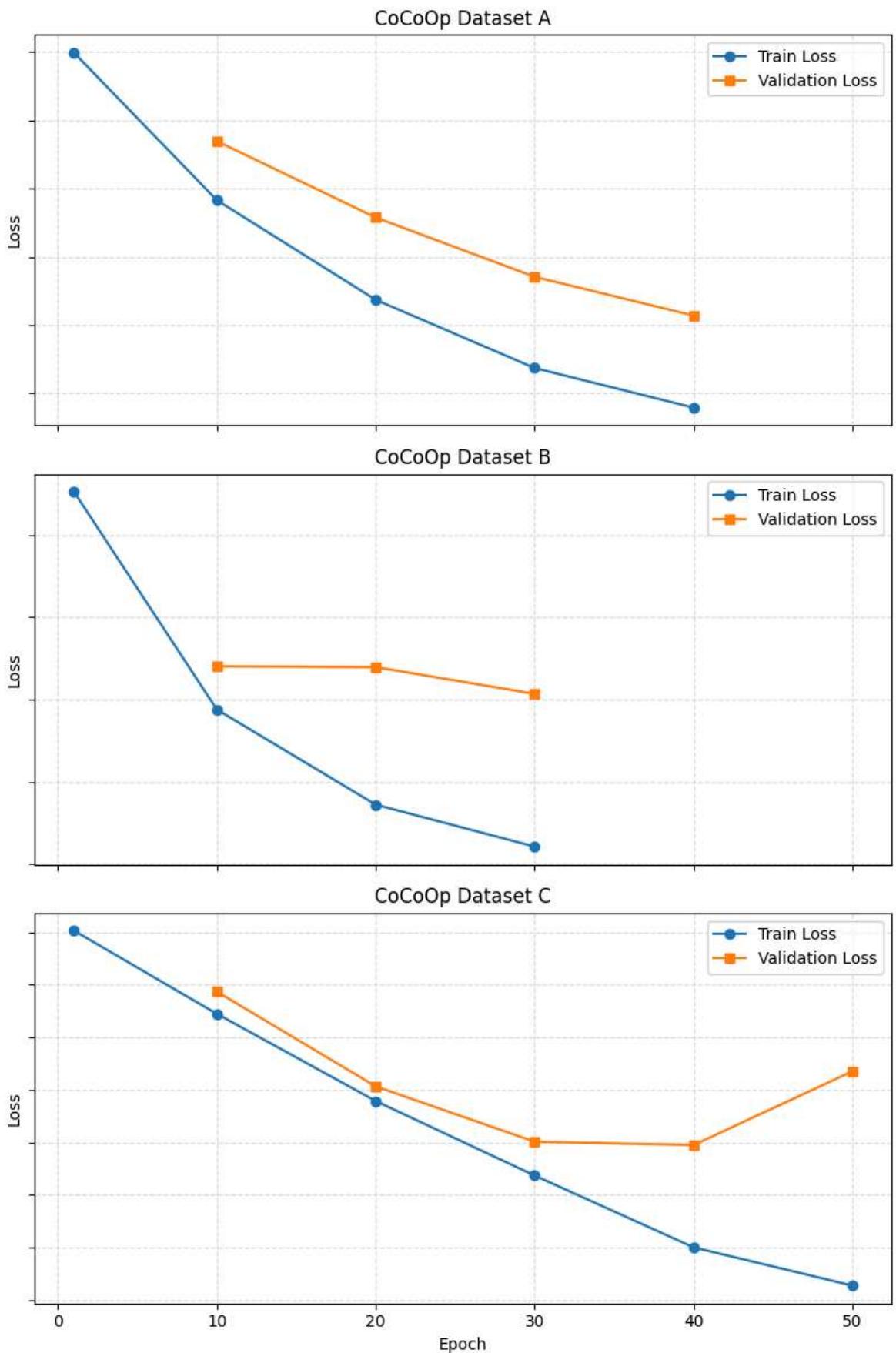
```
Out[23]: dict_keys(['A', 'B', 'C'])
```

```
In [25]: num_plots = len(cocoop_training_curves) or 1
fig, axes = plt.subplots(num_plots, 1, figsize=(8, 4 * num_plots), sharex=True)

if num_plots == 1:
    axes = [axes]

for ax, (dataset_id, df_curve) in zip(axes, sorted(cocoop_training_curves.items())):
    ax.plot(df_curve["epoch"], df_curve["average_loss"], label="Train Loss", marker="o")
    if "validation_average_loss" in df_curve:
        ax.plot(df_curve["epoch"], df_curve["validation_average_loss"], label="Validation Loss", marker="x")
    ax.set_title(f"CoCoOp Dataset {dataset_id}")
    ax.set_ylabel("Loss")
    ax.grid(True, linestyle="--", alpha=0.4)
    ax.legend()
    ax.tick_params(axis="y", labelleft=False)
```

```
axes[-1].set_xlabel("Epoch")
plt.tight_layout()
plt.show()
```



Results

```
In [26]: df = pd.read_csv("../results/task2-2/test/task2_2_results_summary.csv") \
          .sort_values(by=["map", "dataset"], ascending=[False, True])
from IPython.display import HTML
HTML(df.to_html(index=False))
```

Out[26]:

dataset	box_threshold	text_threshold	model_name	map
B	0.7	0.5	cocoop_model_B.pth	0.200625
C	0.5	0.5	cocoop_model_C.pth	0.193288
B	0.3	0.9	cocoop_model_B.pth	0.189167
C	0.5	0.9	cocoop_model_C.pth	0.183432
B	0.3	0.3	cocoop_model_B.pth	0.182438
C	0.5	0.3	cocoop_model_C.pth	0.180519
C	0.5	0.5	cocoop_model_B.pth	0.171406
B	0.7	0.5	cocoop_model_C.pth	0.170490
B	0.7	0.5	cocoop_model_A.pth	0.169786
B	0.3	0.9	cocoop_model_A.pth	0.168036
C	0.5	0.9	cocoop_model_A.pth	0.167895
C	0.5	0.5	cocoop_model_A.pth	0.167879
B	0.3	0.9	cocoop_model_C.pth	0.167710
C	0.5	0.9	cocoop_model_B.pth	0.165589
C	0.3	0.7	cocoop_model_C.pth	0.165458
A	0.3	0.9	cocoop_model_A.pth	0.162356
C	0.7	0.7	cocoop_model_C.pth	0.162167
C	0.5	0.7	cocoop_model_C.pth	0.161990
C	0.5	0.3	cocoop_model_A.pth	0.161161
C	0.5	0.3	cocoop_model_B.pth	0.157126
B	0.7	0.9	cocoop_model_B.pth	0.157101
B	0.3	0.3	cocoop_model_A.pth	0.152733
C	0.5	0.7	cocoop_model_A.pth	0.150706
B	0.3	0.3	cocoop_model_C.pth	0.149958
C	0.5	0.7	cocoop_model_B.pth	0.146226
A	0.7	0.7	cocoop_model_A.pth	0.145923
B	0.7	0.9	cocoop_model_A.pth	0.143535
C	0.3	0.7	cocoop_model_B.pth	0.142910
C	0.3	0.7	cocoop_model_A.pth	0.142703
A	0.3	0.9	cocoop_model_B.pth	0.142176
A	0.5	0.3	cocoop_model_A.pth	0.141282
A	0.3	0.9	cocoop_model_C.pth	0.140688
B	0.7	0.9	cocoop_model_C.pth	0.140207

dataset	box_threshold	text_threshold	model_name	map
C	0.7	0.7	cocoop_model_A.pth	0.140049
A	0.3	0.5	cocoop_model_A.pth	0.138145
C	0.7	0.7	cocoop_model_B.pth	0.136621
C	0.9	0.7	cocoop_model_C.pth	0.134626
A	0.7	0.7	cocoop_model_C.pth	0.133696
C	0.7	0.3	cocoop_model_C.pth	0.133413
A	0.7	0.7	cocoop_model_B.pth	0.133252
A	0.3	0.5	cocoop_model_C.pth	0.123172
C	0.7	0.3	cocoop_model_B.pth	0.121269
C	0.7	0.3	cocoop_model_A.pth	0.121238
C	0.9	0.7	cocoop_model_B.pth	0.120564
A	0.5	0.3	cocoop_model_B.pth	0.120313
C	0.9	0.7	cocoop_model_A.pth	0.120102
A	0.3	0.5	cocoop_model_B.pth	0.119972
B	0.5	0.9	cocoop_model_B.pth	0.119953
A	0.5	0.3	cocoop_model_C.pth	0.118554
C	0.9	0.5	cocoop_model_C.pth	0.117361
C	0.3	0.5	cocoop_model_C.pth	0.115141
A	0.9	0.7	cocoop_model_A.pth	0.111262
B	0.3	0.5	cocoop_model_B.pth	0.109609
A	0.9	0.3	cocoop_model_A.pth	0.108290
A	0.9	0.5	cocoop_model_A.pth	0.106867
A	0.9	0.9	cocoop_model_A.pth	0.105177
C	0.9	0.5	cocoop_model_A.pth	0.102787
B	0.7	0.7	cocoop_model_B.pth	0.102732
C	0.3	0.5	cocoop_model_A.pth	0.101966
C	0.3	0.5	cocoop_model_B.pth	0.101068
A	0.9	0.7	cocoop_model_B.pth	0.100834
C	0.9	0.5	cocoop_model_B.pth	0.100047
B	0.5	0.9	cocoop_model_A.pth	0.099988
A	0.9	0.5	cocoop_model_B.pth	0.098858
A	0.9	0.7	cocoop_model_C.pth	0.098551
A	0.9	0.5	cocoop_model_C.pth	0.098214

dataset	box_threshold	text_threshold	model_name	map
B	0.5	0.9	cocoop_model_C.pth	0.097570
A	0.7	0.3	cocoop_model_A.pth	0.096876
B	0.3	0.5	cocoop_model_C.pth	0.095761
A	0.5	0.5	cocoop_model_A.pth	0.094698
B	0.3	0.5	cocoop_model_A.pth	0.094465
B	0.5	0.7	cocoop_model_B.pth	0.093982
A	0.9	0.3	cocoop_model_B.pth	0.093526
A	0.9	0.3	cocoop_model_C.pth	0.093002
C	0.9	0.9	cocoop_model_C.pth	0.092546
B	0.9	0.3	cocoop_model_B.pth	0.091590
C	0.3	0.3	cocoop_model_C.pth	0.089721
A	0.9	0.9	cocoop_model_B.pth	0.087637
A	0.9	0.9	cocoop_model_C.pth	0.087464
A	0.5	0.5	cocoop_model_C.pth	0.087408
B	0.7	0.7	cocoop_model_A.pth	0.087211
B	0.7	0.7	cocoop_model_C.pth	0.086256
A	0.7	0.3	cocoop_model_B.pth	0.086169
B	0.9	0.3	cocoop_model_C.pth	0.086058
B	0.9	0.3	cocoop_model_A.pth	0.086024
C	0.9	0.9	cocoop_model_B.pth	0.085433
A	0.5	0.5	cocoop_model_B.pth	0.084886
A	0.7	0.3	cocoop_model_C.pth	0.084227
C	0.9	0.9	cocoop_model_A.pth	0.083534
A	0.3	0.3	cocoop_model_A.pth	0.083428
B	0.5	0.7	cocoop_model_C.pth	0.081497
A	0.7	0.5	cocoop_model_A.pth	0.081059
B	0.5	0.7	cocoop_model_A.pth	0.080066
C	0.3	0.3	cocoop_model_A.pth	0.078547
B	0.9	0.5	cocoop_model_B.pth	0.077629
C	0.3	0.3	cocoop_model_B.pth	0.077450
A	0.3	0.3	cocoop_model_C.pth	0.077368
A	0.5	0.9	cocoop_model_A.pth	0.076971
A	0.3	0.3	cocoop_model_B.pth	0.075511

dataset	box_threshold	text_threshold	model_name	map
A	0.7	0.5	cocoop_model_B.pth	0.073434
A	0.7	0.5	cocoop_model_C.pth	0.073069
B	0.9	0.5	cocoop_model_C.pth	0.070662
A	0.7	0.9	cocoop_model_A.pth	0.070289
B	0.9	0.5	cocoop_model_A.pth	0.068850
A	0.5	0.9	cocoop_model_C.pth	0.068613
A	0.5	0.9	cocoop_model_B.pth	0.067480
B	0.5	0.3	cocoop_model_B.pth	0.062895
A	0.5	0.7	cocoop_model_A.pth	0.062019
C	0.7	0.9	cocoop_model_C.pth	0.061282
A	0.7	0.9	cocoop_model_B.pth	0.060895
A	0.7	0.9	cocoop_model_C.pth	0.059879
C	0.7	0.5	cocoop_model_C.pth	0.058508
B	0.5	0.3	cocoop_model_C.pth	0.055271
B	0.5	0.3	cocoop_model_A.pth	0.054827
C	0.7	0.9	cocoop_model_B.pth	0.054229
C	0.7	0.9	cocoop_model_A.pth	0.054218
A	0.3	0.7	cocoop_model_A.pth	0.052821
A	0.5	0.7	cocoop_model_B.pth	0.052539
A	0.5	0.7	cocoop_model_C.pth	0.050663
C	0.7	0.5	cocoop_model_B.pth	0.049997
C	0.7	0.5	cocoop_model_A.pth	0.049179
B	0.5	0.5	cocoop_model_B.pth	0.047734
B	0.9	0.9	cocoop_model_B.pth	0.046516
C	0.9	0.3	cocoop_model_C.pth	0.046328
A	0.3	0.7	cocoop_model_C.pth	0.045231
A	0.3	0.7	cocoop_model_B.pth	0.043908
B	0.5	0.5	cocoop_model_C.pth	0.043470
B	0.5	0.5	cocoop_model_A.pth	0.042937
B	0.9	0.7	cocoop_model_B.pth	0.042035
B	0.7	0.3	cocoop_model_B.pth	0.041581
C	0.9	0.3	cocoop_model_B.pth	0.041057
C	0.3	0.9	cocoop_model_C.pth	0.040785

dataset	box_threshold	text_threshold	model_name	map
B	0.9	0.9	cocoop_model_C.pth	0.039634
B	0.9	0.9	cocoop_model_A.pth	0.039574
C	0.9	0.3	cocoop_model_A.pth	0.039491
B	0.3	0.7	cocoop_model_B.pth	0.036348
B	0.9	0.7	cocoop_model_C.pth	0.036032
C	0.3	0.9	cocoop_model_B.pth	0.035981
B	0.9	0.7	cocoop_model_A.pth	0.035760
B	0.7	0.3	cocoop_model_A.pth	0.035159
B	0.7	0.3	cocoop_model_C.pth	0.034849
C	0.3	0.9	cocoop_model_A.pth	0.034618
B	0.3	0.7	cocoop_model_A.pth	0.032218
B	0.3	0.7	cocoop_model_C.pth	0.032039

Task 3 Fix Match

```
In [27]: task3_train_path = Path("../results/task3/train/dataset_B-C.json")
with open(task3_train_path, "r", encoding="utf-8") as f:
    bc_payload = json.load(f)
bc_history = pd.DataFrame(bc_payload.get("history", []))

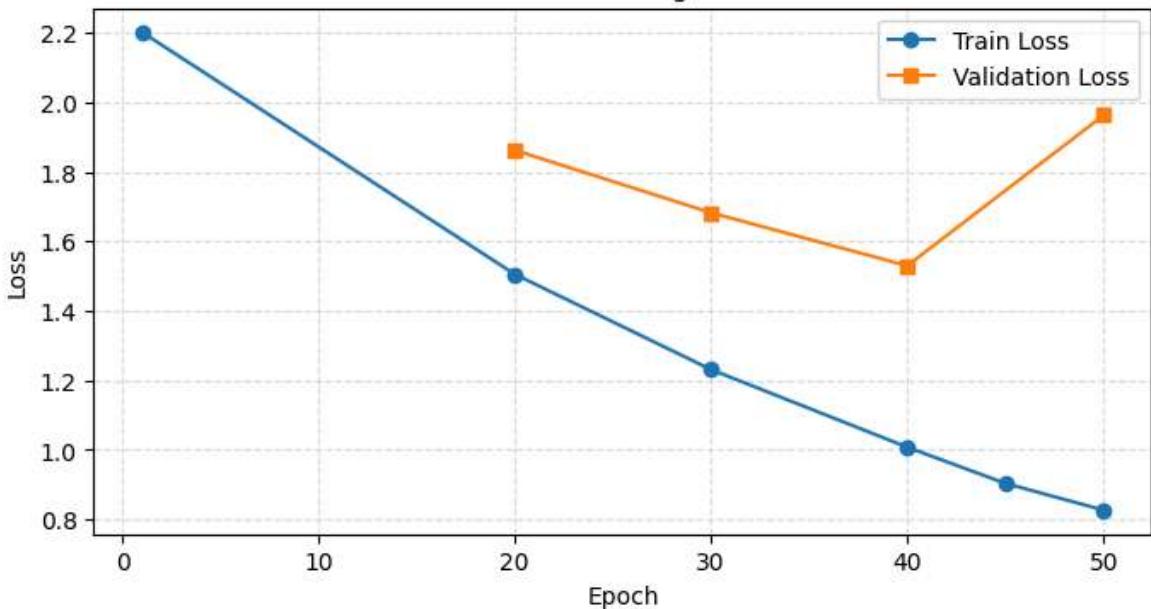
fig, ax = plt.subplots(figsize=(8, 4))
ax.plot(bc_history["epoch"], bc_history["average_loss"], label="Train Loss", marker="o")

if "validation_average_loss" in bc_history:
    val_mask = bc_history["validation_average_loss"].notna()
    if val_mask.any():
        ax.plot(
            bc_history.loc[val_mask, "epoch"],
            bc_history.loc[val_mask, "validation_average_loss"],
            label="Validation Loss",
            marker="s",
        )

ax.set_title("Task 3 FixMatch Training - Dataset B→C")
ax.set_xlabel("Epoch")
ax.set_ylabel("Loss")
ax.grid(True, linestyle="--", alpha=0.4)
ax.legend()
plt.show()

bc_history
```

Task 3 FixMatch Training - Dataset B→C



Out[27]:

	epoch	average_loss	validation_average_loss
0	1	2.2022	NaN
1	20	1.5054	1.8624
2	30	1.2312	1.6824
3	40	1.0078	1.5299
4	45	0.9037	NaN
5	50	0.8267	1.9628

Results

In [29]:

```
df = pd.read_csv("../results/task3/test/task3_results_summary.csv") \
    .sort_values(by=["map", "dataset_pair"], ascending=[False, True])
from IPython.display import HTML
HTML(df.to_html(index=False))
```

Out[29]:

dataset_pair	box_threshold	text_threshold	model_name	map
B-C	0.7	0.3	fm_coop_model_B-C.pth	0.284454
B-C	0.5	0.9	fm_coop_model_B-C.pth	0.247580
B-C	0.9	0.3	fm_coop_model_B-C.pth	0.227924
B-C	0.9	0.5	fm_coop_model_B-C.pth	0.185833
B-C	0.7	0.7	fm_coop_model_B-C.pth	0.171641
B-C	0.7	0.9	fm_coop_model_B-C.pth	0.165355
B-C	0.3	0.5	fm_coop_model_B-C.pth	0.139663
B-C	0.9	0.9	fm_coop_model_B-C.pth	0.124546
B-C	0.7	0.5	fm_coop_model_B-C.pth	0.112539
B-C	0.5	0.3	fm_coop_model_B-C.pth	0.108558
B-C	0.3	0.3	fm_coop_model_B-C.pth	0.107744
B-C	0.3	0.7	fm_coop_model_B-C.pth	0.095333
B-C	0.5	0.7	fm_coop_model_B-C.pth	0.082718
B-C	0.3	0.9	fm_coop_model_B-C.pth	0.082668
B-C	0.9	0.7	fm_coop_model_B-C.pth	0.080184
B-C	0.5	0.5	fm_coop_model_B-C.pth	0.079162

In []:

In []:

In []:

In []: