Segmentation Task

- **Dataset Choice:**
  I chose the **ModaNet dataset**, which covers **13 clothing classes** and contains a total of **52,225 images**. I split the dataset into **60% training, 25% validation, and 15% test sets**. ModaNet replicates real-world street-fashion images annotated with **polygon masks**, providing rich diversity that makes models trained on it more **robust to variations in lighting, pose, background, and occlusion**. However, the dataset exhibits significant **class imbalances**, which presented challenges during training.

- **Model Architecture:**
  I used a standard U-Net architecture implemented through the `segmentation_models_pytorch (SMP)` library**. U-Net was chosen for its** simplicity and strong performance in pixel-level segmentation tasks. The architecture follows the classic encoder–decoder design, but SMP enhances it by integrating optional attention mechanisms in the decoder, improving the model's ability to focus on relevant spatial regions.

  Only the final layer was modified to output 13 segmentation classes corresponding to the dataset categories. U-Net proved easy to train and effective when combined with specialized loss functions such as Focal Loss, which addresses class imbalance by focusing more on difficult or underrepresented classes.

- **Loss Function:**
  Initially, I trained the model using **categorical cross-entropy loss**, but classes with **fewer pixel representations** consistently scored 0 on all metrics, indicating the model failed to learn these minority classes. To address this, I incorporated **Weighted Dice Loss** and **Focal Loss**, where the weights were calculated by looping through the dataset and computing each class's pixel percentage. Additionally, to improve **Intersection over Union (IoU)** performance, I added **Lovász-Softmax loss**. I combined these three losses into a **single weighted loss function**, with each component scaled according to the class pixel distribution.

● Performance Analysis and Evaluation Metrics

The model's performance was evaluated using standard segmentation metrics, including **IoU (Intersection over Union)**, **Precision**, **Recall**, and **F1-Score**.

In the initial trials, the model showed **imbalanced performance** across classes — some **minor or underrepresented classes** (such as small accessories) achieved **very low or even zero scores** across most metrics. This was mainly due to the **class imbalance** in the dataset, where dominant classes like background or large clothing items overshadowed smaller ones.

To address this, **Focal Loss**, **Weighted Cross-Entropy**, and **Dice Loss** were incorporated into the training process. The **Focal Loss** helped the model focus on difficult and less frequent classes, while the **Dice and IoU-based losses** improved the overlap accuracy of predicted masks.

After integrating these loss functions, the model achieved **significant performance improvements**, especially for previously underperforming classes. The segmentation became more balanced and consistent across all categories, indicating better generalization and more stable learning dynamics

Run All   Code

```
Testing: 100%|██████████| 489/490 [01:56<00:00,  5.38batch/s, Memory cleaned at batch 480]/tm
rning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda'
  with torch.cuda.amp.autocast():
Testing: 100%|██████████| 490/490 [01:56<00:00,  4.20batch/s, Memory cleaned at batch 480]
```

	Class	Precision	Recall	F1-Score	IoU	Support
0	bag	0.9840	0.9904	0.9872	0.9747	425867341
1	belt	0.7936	0.7046	0.7465	0.5955	4586983
2	boots	0.3816	0.0263	0.0492	0.0252	239463
3	dress	0.7216	0.5762	0.6408	0.4714	1632219
4	footwear	0.8004	0.8226	0.8114	0.6826	13535347
5	headwear	0.7045	0.7058	0.7051	0.5446	3454322
6	outer	0.8262	0.2844	0.4232	0.2684	633474
7	pants	0.8109	0.8090	0.8100	0.6807	16125233
8	scarf-tie	0.8709	0.8743	0.8726	0.7740	13105055
9	shorts	0.0000	0.0000	0.0000	0.0000	1164037
10	skirt	0.8117	0.7585	0.7842	0.6450	1836099
11	sunglasses	0.8145	0.8002	0.8073	0.6769	8755341
12	top	0.9223	0.0406	0.0778	0.0404	245923
13	Macro Avg	0.7263	0.5687	0.5935	0.4907	491180837

\+ Code   + Markdown

Trial one image

You can see that class shorts got 0 in all metrics the one with low support imbalance is also getting low metrics

Trial 2 with the loss functions focal,weighted dice and `lovasz_softmax`
All combined with a weights to take into account to whole loss

Run All   Code   Draft

```
Testing: 100%|██████████| 490/490 [01:55<00:00,  4.24batch/s, Memory cle
```

	Class	Precision	Recall	F1-Score	IoU	Support
0	bag	0.9975	0.9482	0.9722	0.9459	423879700
1	belt	0.4698	0.8768	0.6118	0.4407	4600335
2	boots	0.2273	0.7859	0.3527	0.2141	314170
3	dress	0.4671	0.7641	0.5798	0.4083	1632198
4	footwear	0.7108	0.7992	0.7524	0.6031	13730768
5	headwear	0.4528	0.8796	0.5979	0.4264	3455955
6	outer	0.3482	0.8799	0.4989	0.3324	633263
7	pants	0.6610	0.7894	0.7195	0.5619	16798596
8	scarf-tie	0.7691	0.8828	0.8221	0.6979	13124271
9	shorts	0.2670	0.7021	0.3868	0.2398	1342276
10	skirt	0.6215	0.8284	0.7102	0.5506	1950831
11	sunglasses	0.7326	0.8236	0.7754	0.6332	8831439
12	top	0.6738	0.8571	0.7545	0.6057	245517
13	Macro Avg	0.5691	0.8321	0.6565	0.5123	490539319

+ Code   + Markdown

a. Strengths of the System:

- The model is **easy to train and adapt** to different datasets.
- It **accurately detects and segments most clothing items** across the 13 defined classes.
- The system shows **strong performance on single-person images**, especially when clothing is clear and not overlapped with other objects.
- The model effectively captures **fine clothing details** like dress or pants boundaries when the background is distinct.

---

b. Drawbacks of the System:

- **Data imbalance** affected performance—some classes remained underrepresented, leading to weaker predictions despite using advanced loss functions (e.g., Focal Loss) and augmentations.
- Model over segment to the body

- The model **sometimes misclassifies hair as headwear** or incorrectly segments when colors and textures are similar.

- It **struggles in complex scenes** with multiple people or overlapping clothing items.

- The system primarily fits **women's clothing classes**, which limits its ability to generalize to a wider range of apparel. While basic augmentations were applied and propalistic to rare which made gpu training hard , performance could improve with **more diverse and probabilistic augmentations** and a **larger, balanced dataset**. The model occasionally **misses small clothing parts** or **over-segments body regions**, indicating the need for **more data** and **class-aware training strategies**. Future improvements could include expanding the dataset to cover **broader clothing categories**, adopting **more advanced model architectures**, and training on **higher-end GPUs** to enhance segmentation accuracy and robustness

- A deeper exploration and tuning of the **focal loss parameters (γ and α)** would likely have improved the results. **The drop in precision** occurred because the model became more focused on minority classes, but on the positive side, classes that previously had low or zero metrics now appeared and achieved measurable scores