Face Alignment Experiments Report, Покумин Георгий

Краткое содержание экспериментов

Список проведенных экспериментов и результаты:

Эксперимент	AUC (300W)	AUC (Menpo)
EfficientNet + Heatmap + BCE	0.9493	0.9829
EfficientNet + Heatmap + MSE	0.9508	0.9827
ConvNeXt + Heatmap + BCE	0.9400	0.9826
ConvNeXt + Heatmap + MSE	0.9500	0.9820
EfficientNet + Regression + WIng Loss	0.9359	0.9815
EfficientNet + Regression + Adaptive WIng Loss	0.9346	0.9798
EfficientNet + Regression + MSE	0.9284	0.9756
ConvNeXt + Regression + Adaptive Wing Loss	0.9148	0.9719
EfficientNet + Heatmap + Focal Loss	0.9413	0.9712
ConvNeXt + Regression + WIng Loss	0.9175	0.9709
ConvNeXt + Heatmap + Focal Loss	0.9355	0.9679
ConvNeXt + Regression + MSE	0.9087	0.9595
DLIB	-	0.9383

Подробное описание методики и реализации

Данные

Для обучения, владиации и тестирования использовались два датасета - 300W и Menpo. При обучении и валидации они перемешивались, тестирование проходило на каждом независимо. Для валидации

использовалось 20% совмещенной тренировочной выборки 300W и Menpo, остальное - для обучения.

На этапе предобработки изображения считывались с помощью библиотеки PIL и переводились в формат RGB. Далее выполнялась детекция лиц через детектор dlib. При детектировании лица с помощью dlib в коде существует две опции: использование предварительно рассчитанных боксов (эту задачу выполняет скрипт precompute_boxes.py), либо расчет на ходу. При отсутствии результата использовался весь кадр. Чтобы избежать пропусков некоторых ключевых точек, прямоугольник корректировался, так чтобы включать все, например, левый верхний угол - минимум из угла детектора и и верхней левой точки. После этого выполнялось обрезание изображения по области лица с небольшим расширением (10% от размеров кропа), чтобы не терять контекст, и пересчёт координат ключевых точек в систему координат внутри обрезанного изображения. Затем изображения приводились к единому размеру (224x224) и проходили серию аугментаций, включающих изменение яркости, контрастности, небольшие повороты, сдвиги и масштабирование, после чего нормализовались по статистикам ImageNet. Ключевые точки также проходили нормализацию (к диапазону [0, 1]) относительно полученного изображения для лучшей сходимости.

Обучение

Для обучения применялись две архитектуры нейронных сетей - EfficientNet-B0 и ConvNeXtV2-Nano, обе реализованные через библиотеку timm и использовавшиеся в качестве извлекающих признаков блоков (backbone). Поверх них добавлялись два типа выходных голов: регрессионная и тепловая (heatmap). В первом случае сеть напрямую предсказывала нормализованные координаты ключевых точек, а во втором - генерировала карты активаций, где каждая точка представлялась гауссовым пятном. Из карт затем восстанавливались координаты с помощью операции softmax. Регрессионная голова состояла из двух линейных слоев, соединенных функцией активации ReLu и нормализацией Dropout. Heatmap голова состояла из двух блоков свертка - батч-норма - ReLu и последнего

сверточного слоя. Везде брали последнюю карту признаков из выхода backbone.

Обучение проводилось с использованием оптимизатора AdamW и косинусного расписания скорости обучения. Размер батча составлял 64, а обучение продолжалось 40 эпох. Для регрессионной головы применялись функции потерь MSE, WingLoss и AdaptiveWingLoss, а для тепловых карт - MSE (здесь выходы модели дополнительно проходили через sigmoid), фокальная потеря и кросс-энтропия.

Результаты

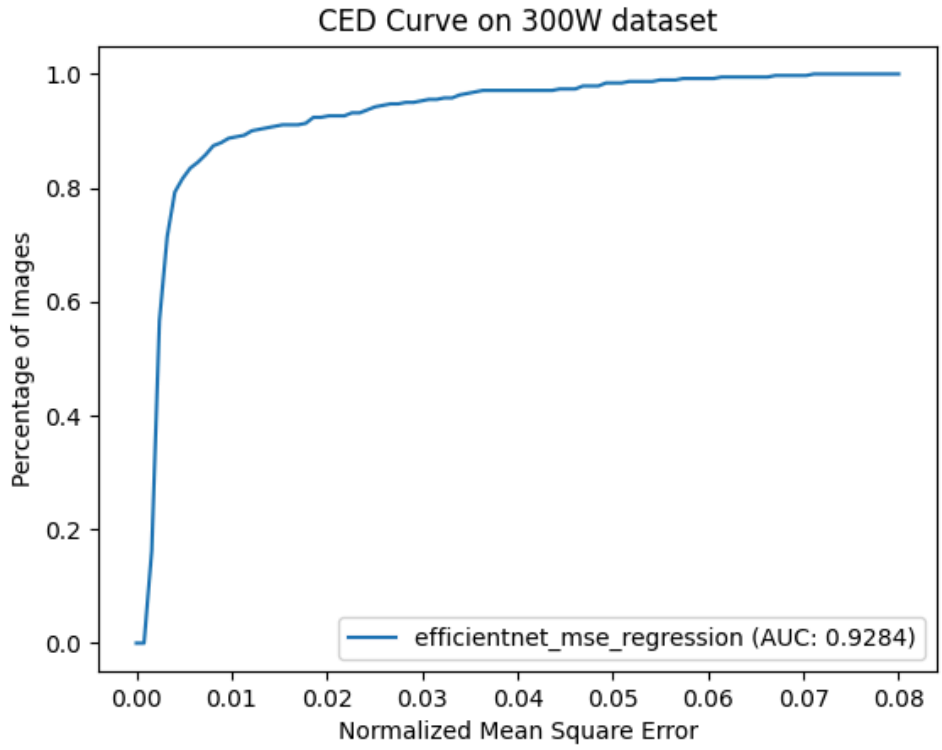
Результаты экспериментов показали, что подход с тепловыми картами превосходит прямую регрессию координат по точности локализации - они показали высочайшие значения метрик на обоих датасетах, что соответствует ожиданиям, так как подход на основе тепловых карт более устойчив к перекрытиям, масштабированию и шумам за счет моделирования плотности расположения точки. Также в среднем архитектуры на основе EfficientNet показывают лучшее качество. Регрессионный подход показывает конкурентное качество при простоте реализации, особенно при использовании Wing/AdaptiveWing функций потерь (показали себя лучше стандартной MSE). Также все модели превзошли dlib по качеству на датасете Menpo со значительным отрывом.

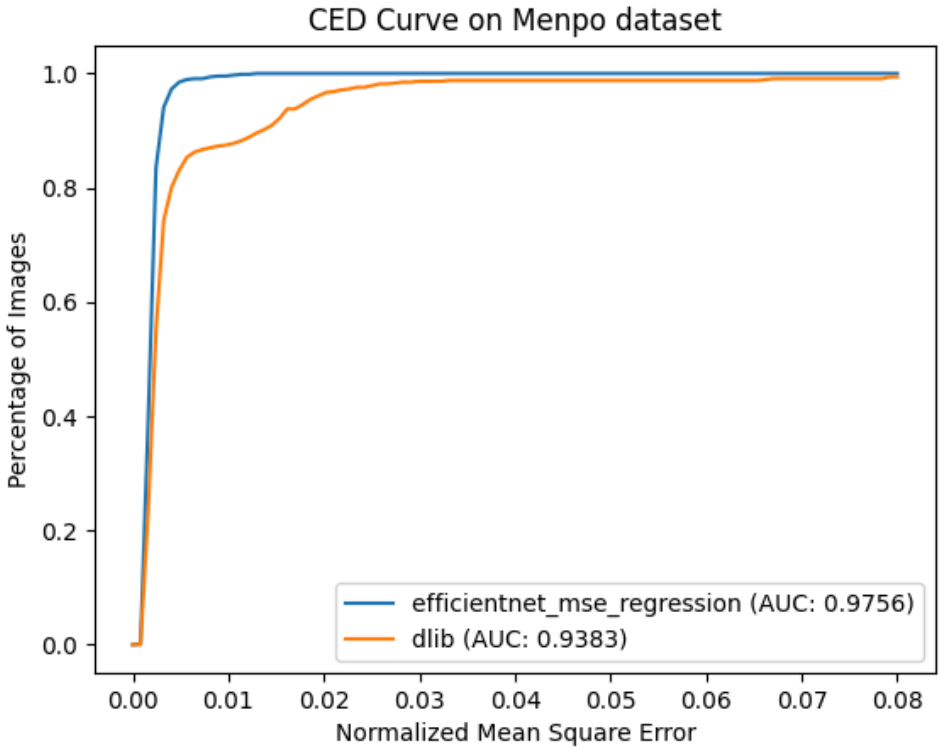
Итоговая лучшая модель основана на EfficientNet-B0 с тепловой головой и функцией потерь MSE. Она обладает около пяти миллионами параметров, что делает её достаточно компактной при высоком качестве предсказаний. Такая комбинация архитектуры и функции потерь обеспечивает хорошее соотношение между точностью, скоростью и устойчивостью модели к различным условиям освещения и положению лица в кадре.

Таким образом, по результатам экспериментов можно сделать несколько выводов. Во-первых, использование тепловых карт для предсказания координат дает явное преимущество по качеству и

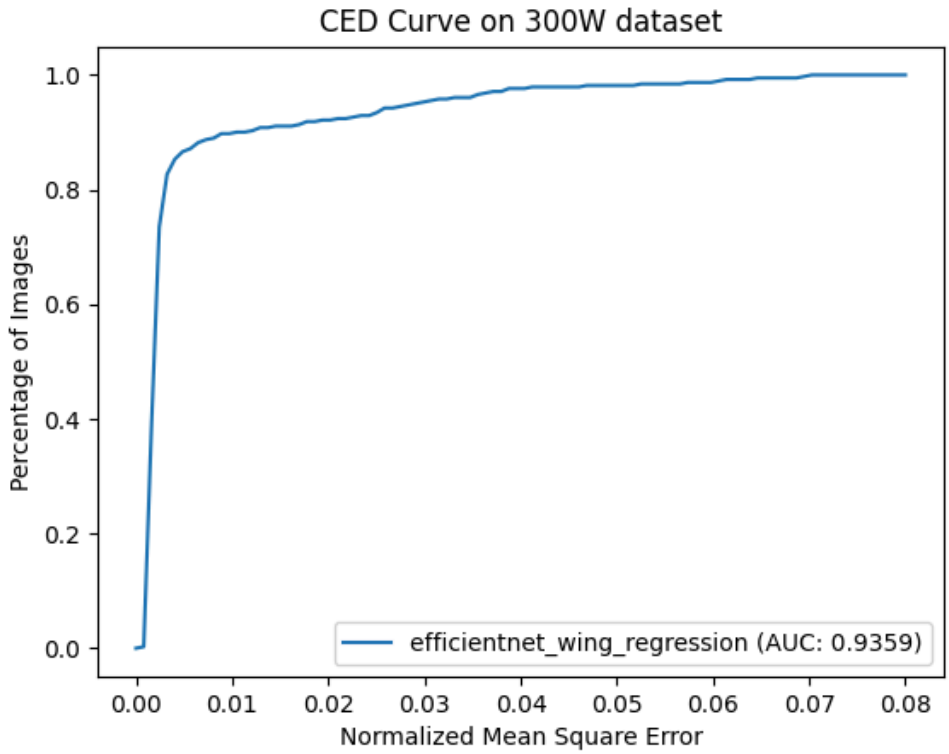
стабильности. Во-вторых, EfficientNet-B0 оказался оптимальным выбором в соотношении производительности и вычислительных затрат. Все CED кривые и значения AUC можно посмотреть ниже.

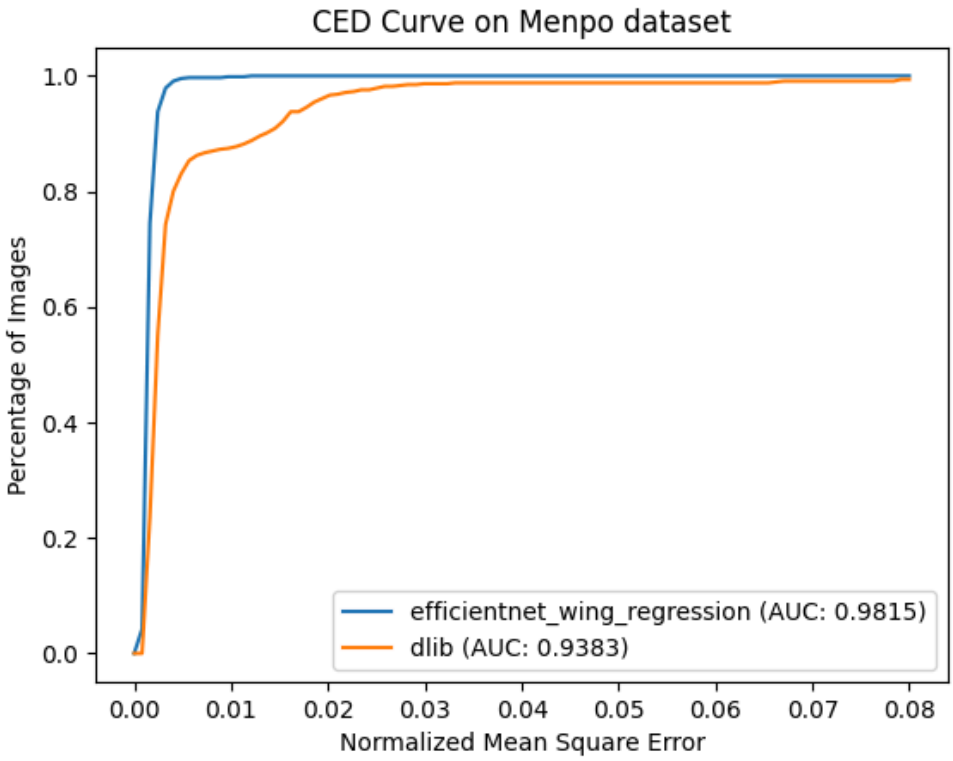
EfficientNet + MSE + Regression



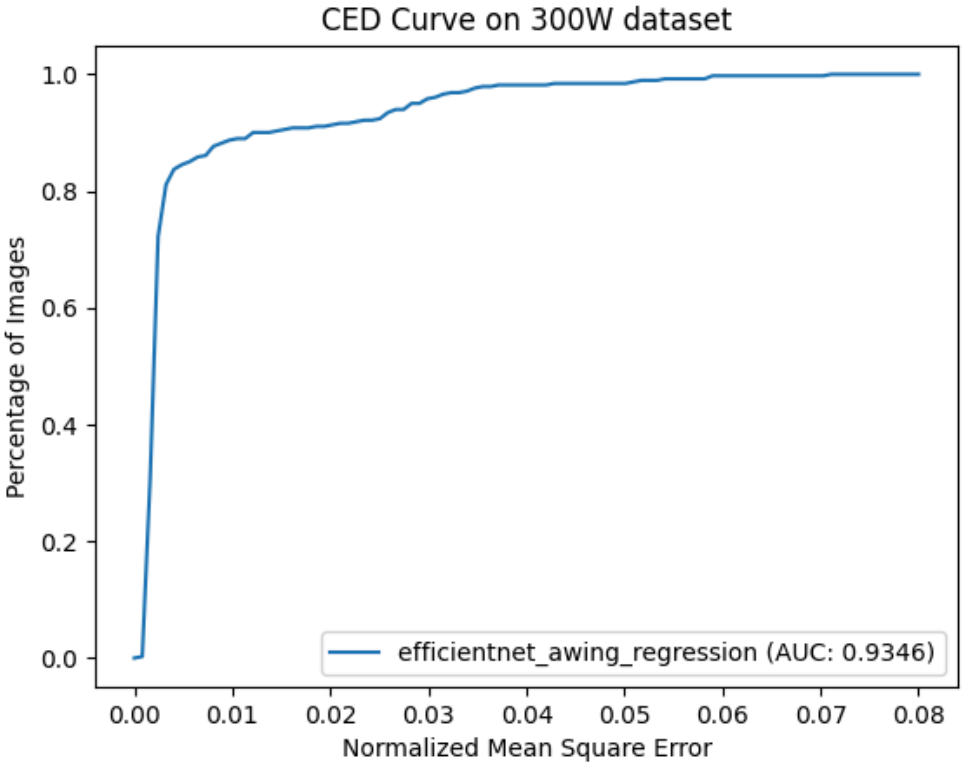


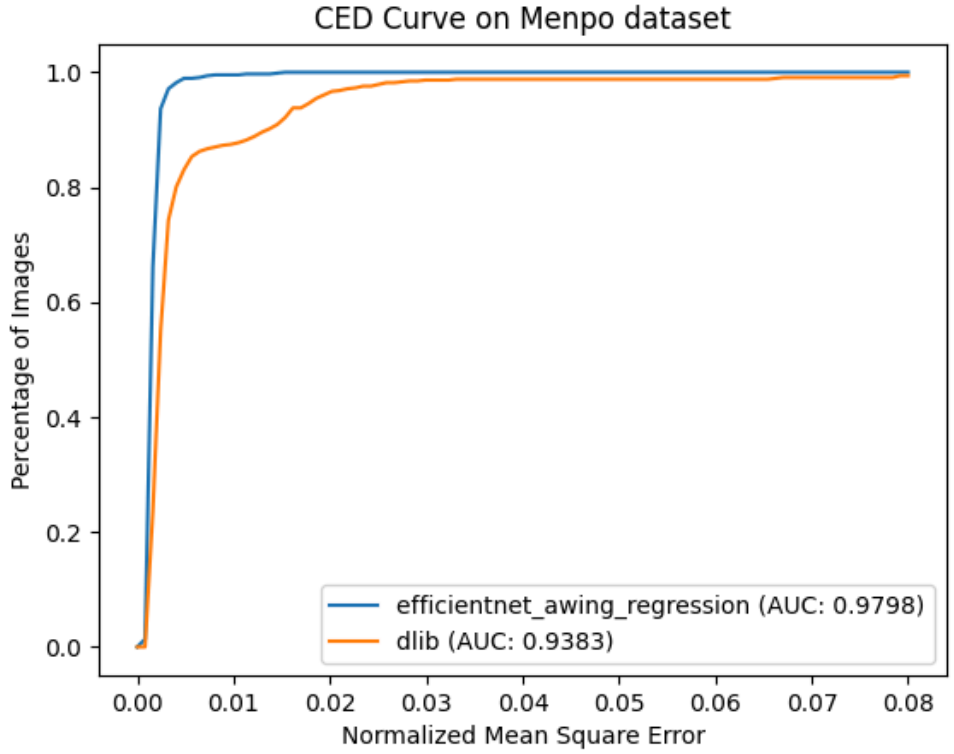
EfficientNet + Wing Loss + Regression



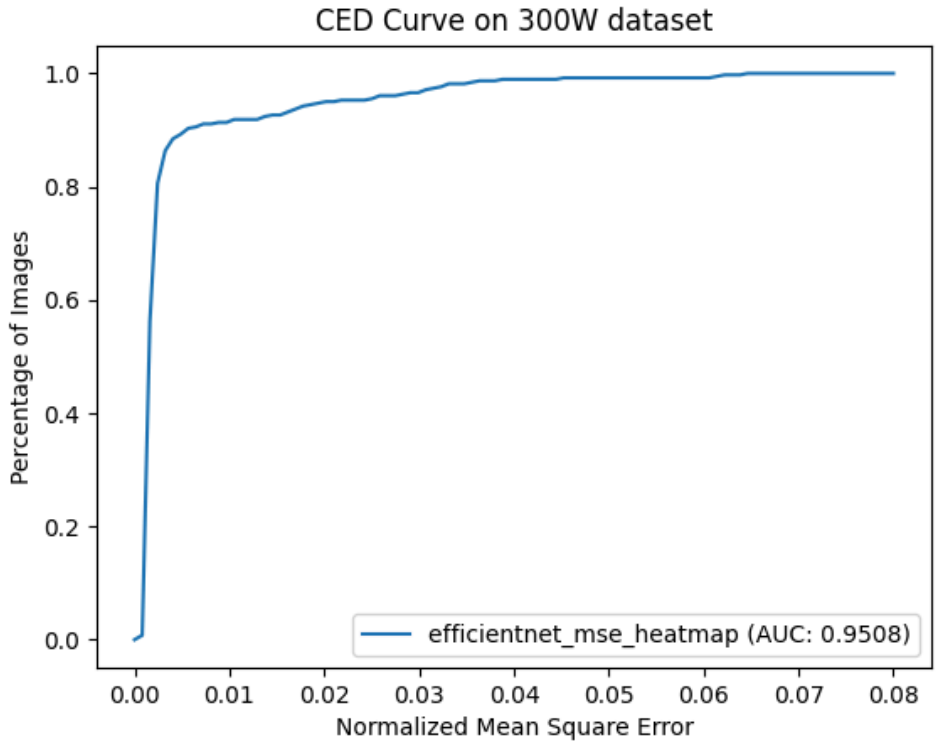


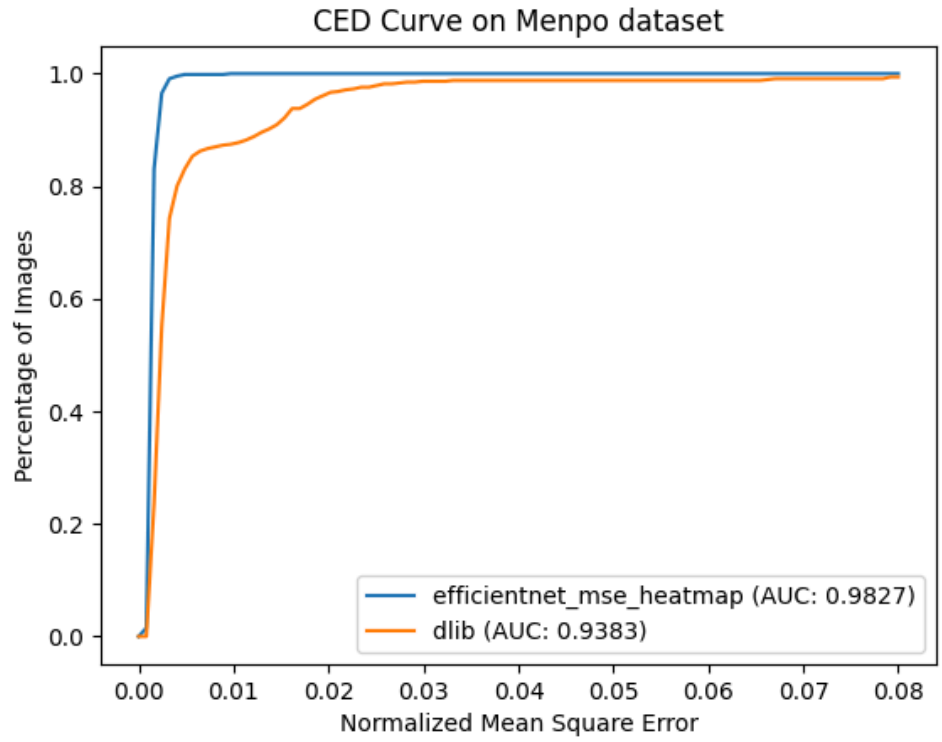
EfficientNet + AdaptiveWing + Regression



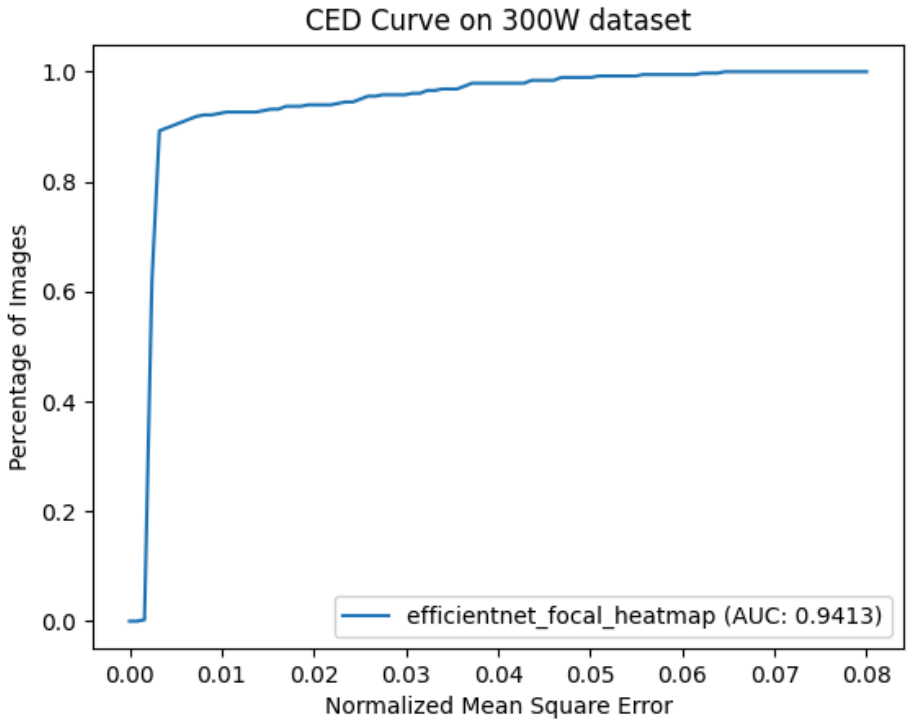


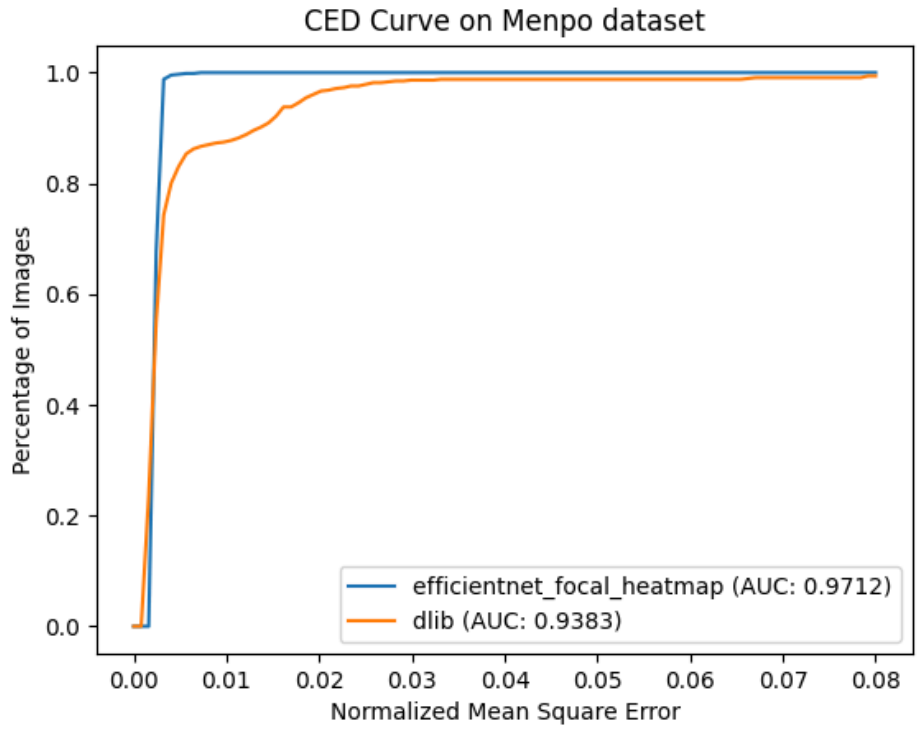
EfficientNet + MSE + Heatmap



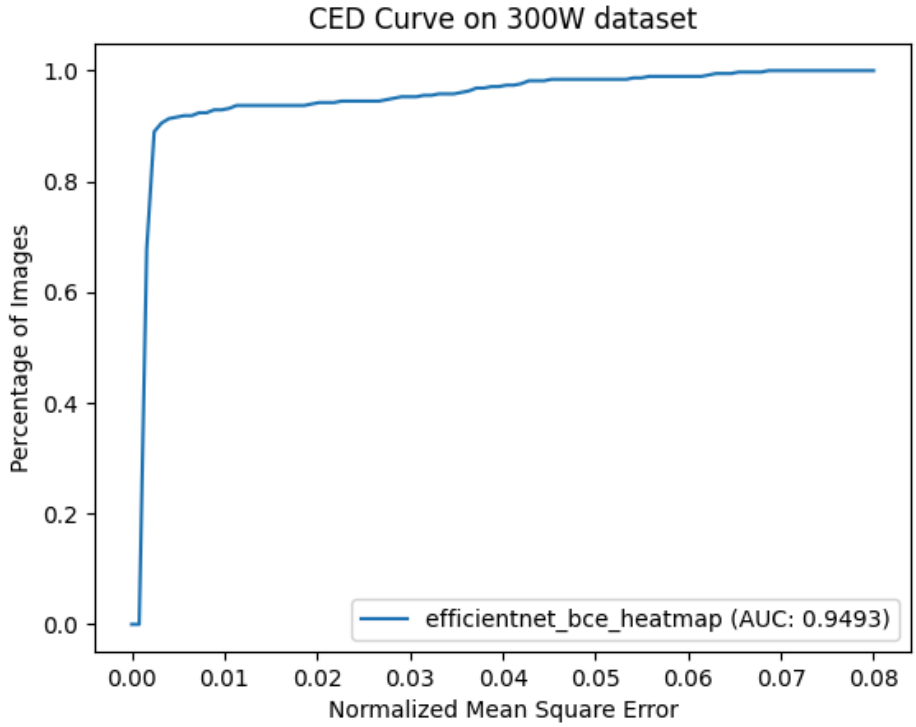


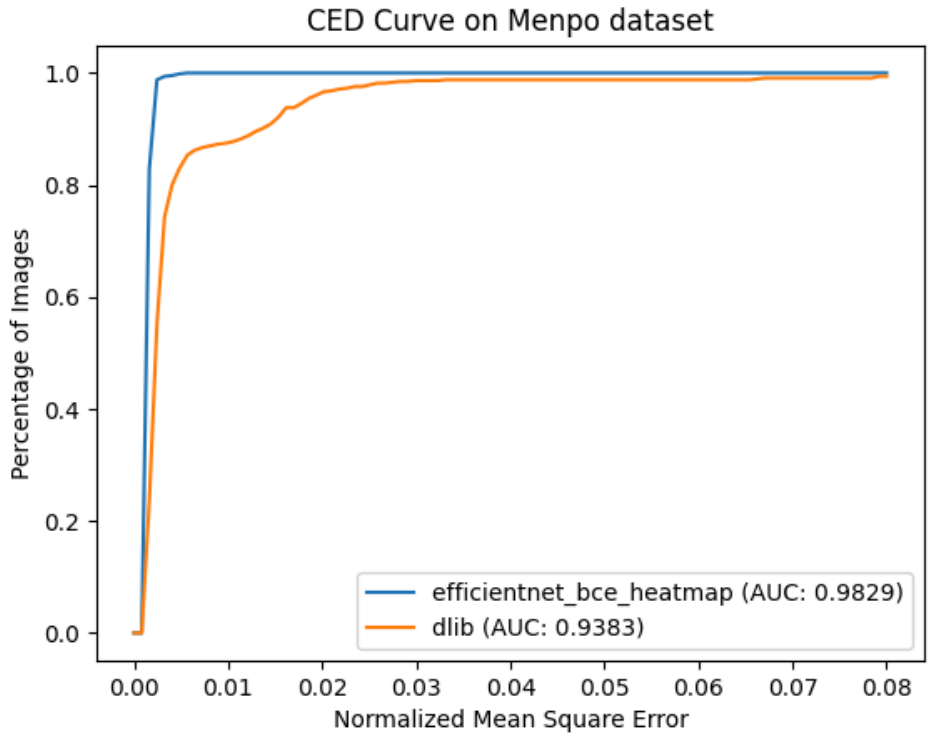
EfficientNet + FocalLoss + Heatmap



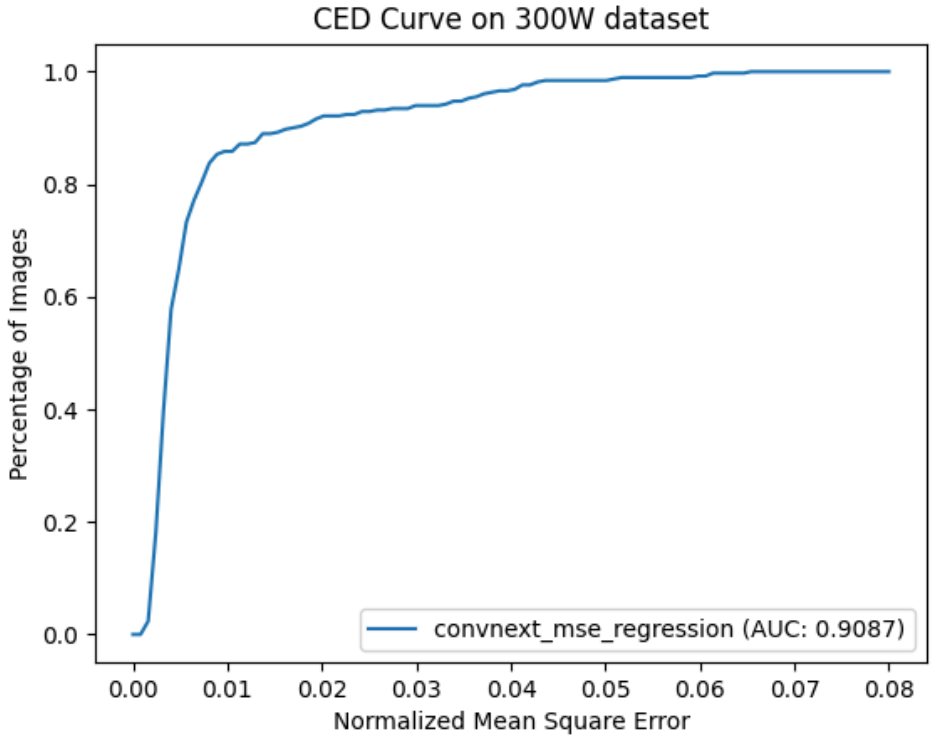


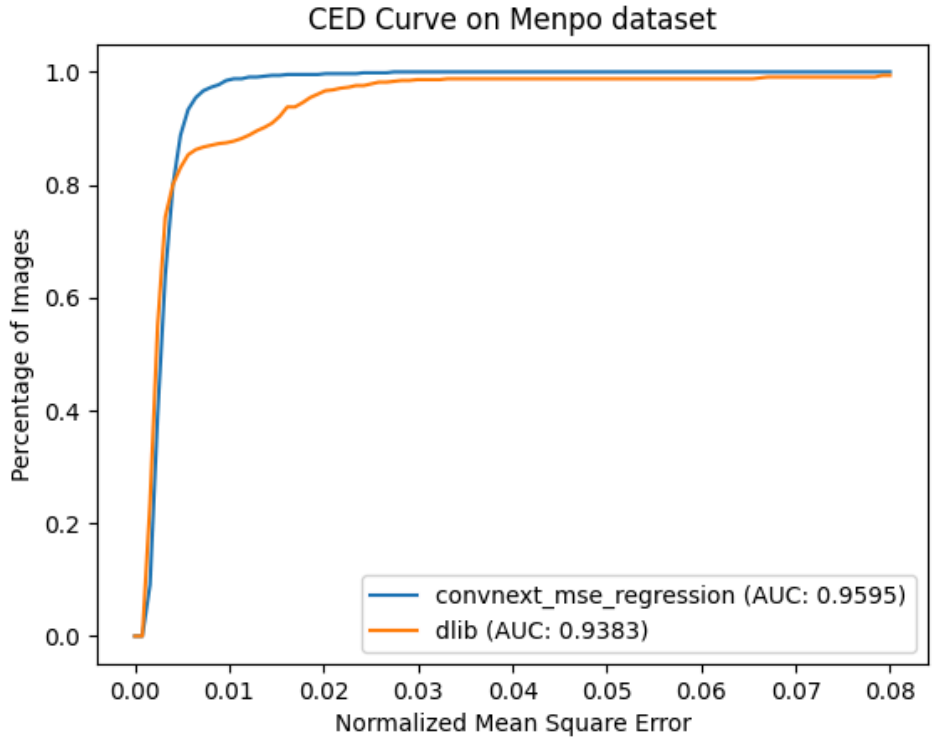
EfficientNet + BCE + Heatmap



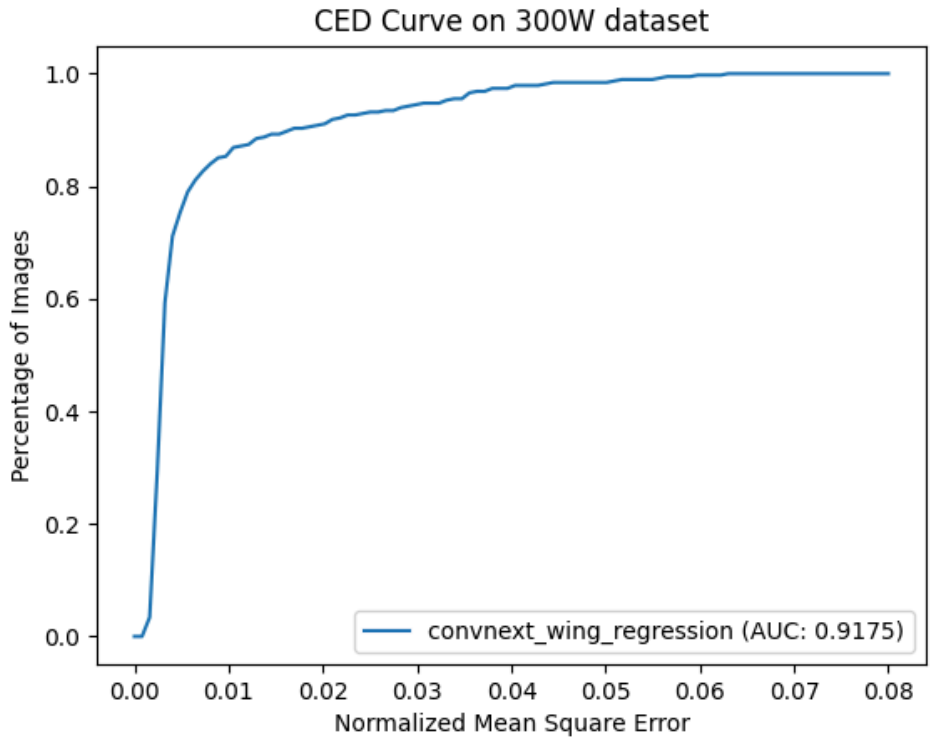


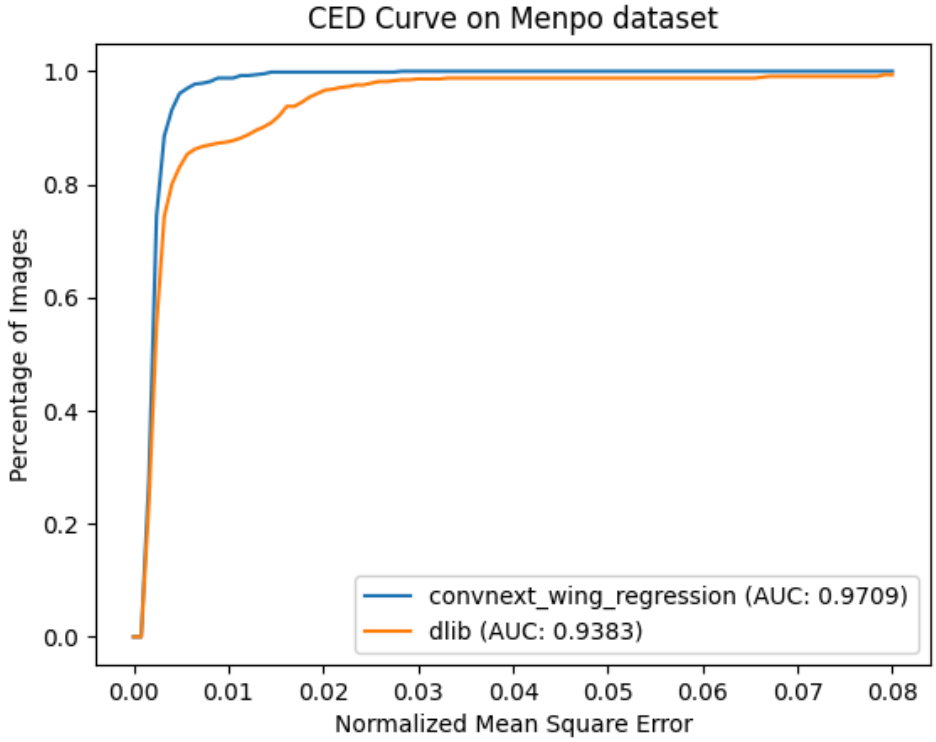
ConvNeXt + MSE + Regression



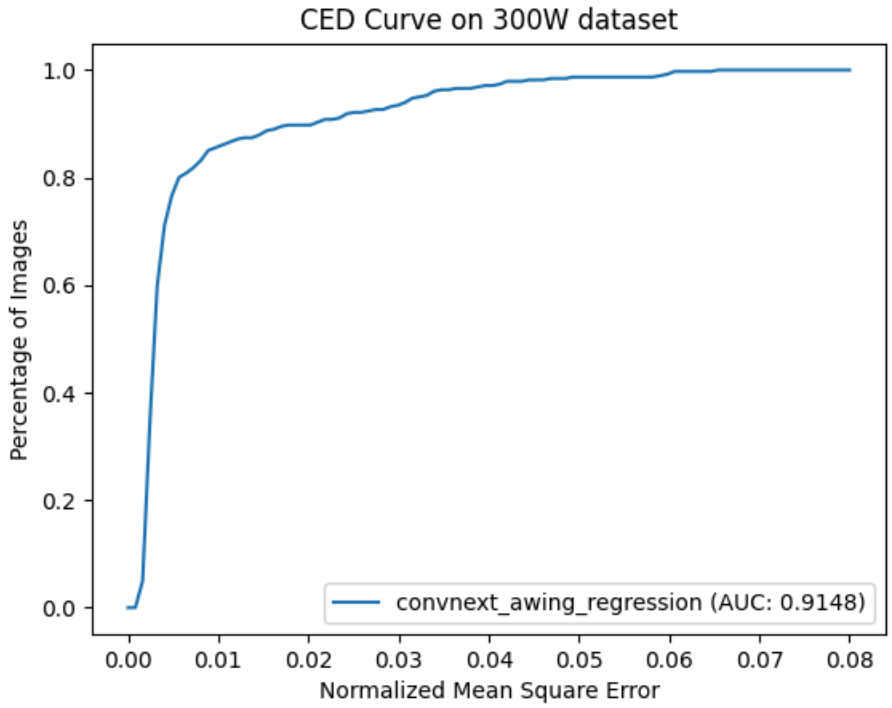


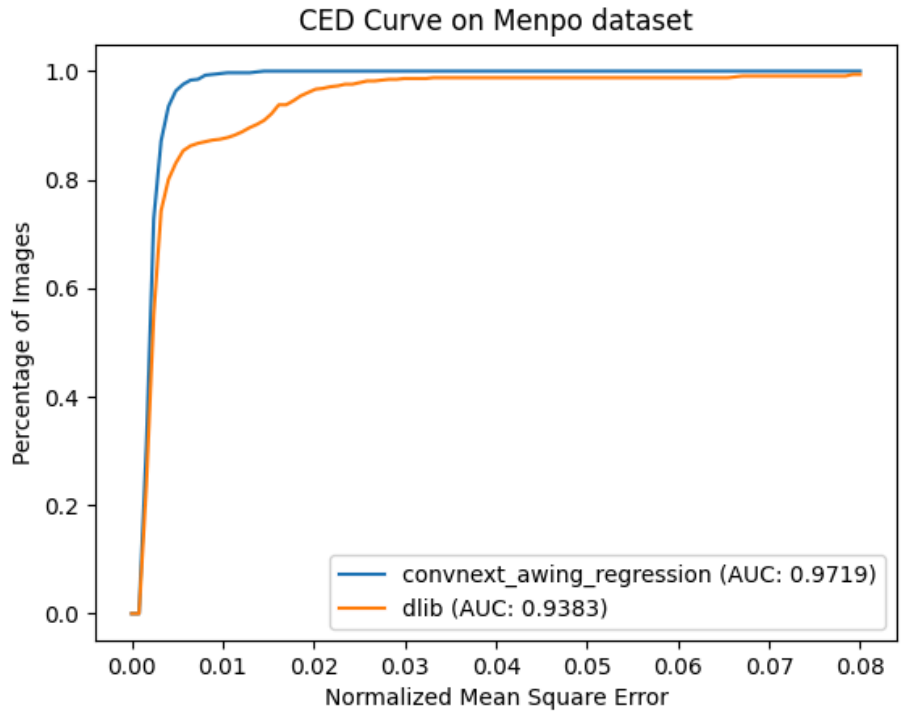
ConvNeXt + Wing Loss + Regression



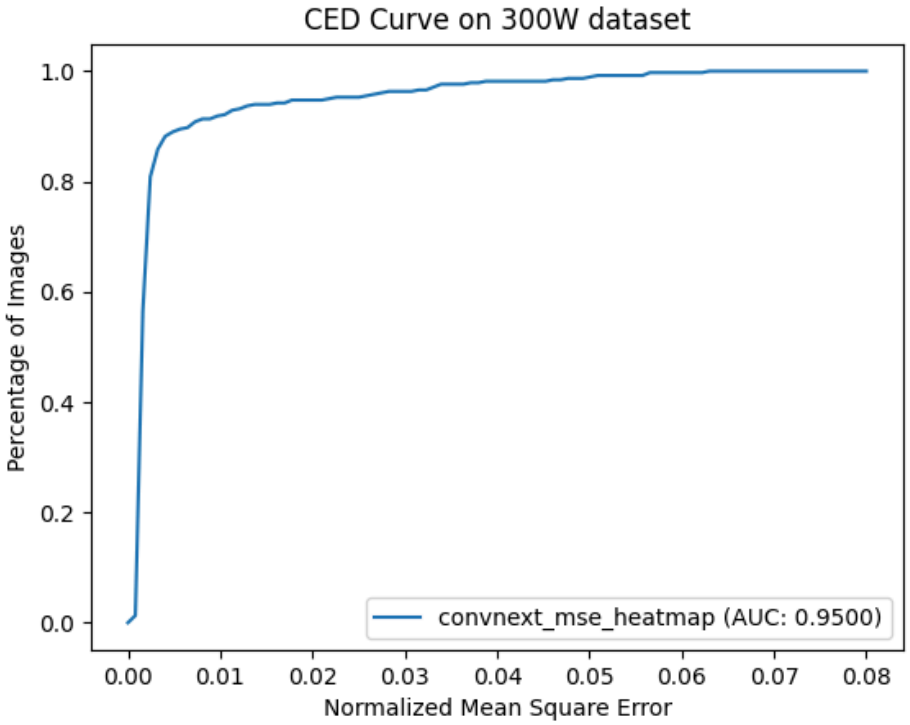


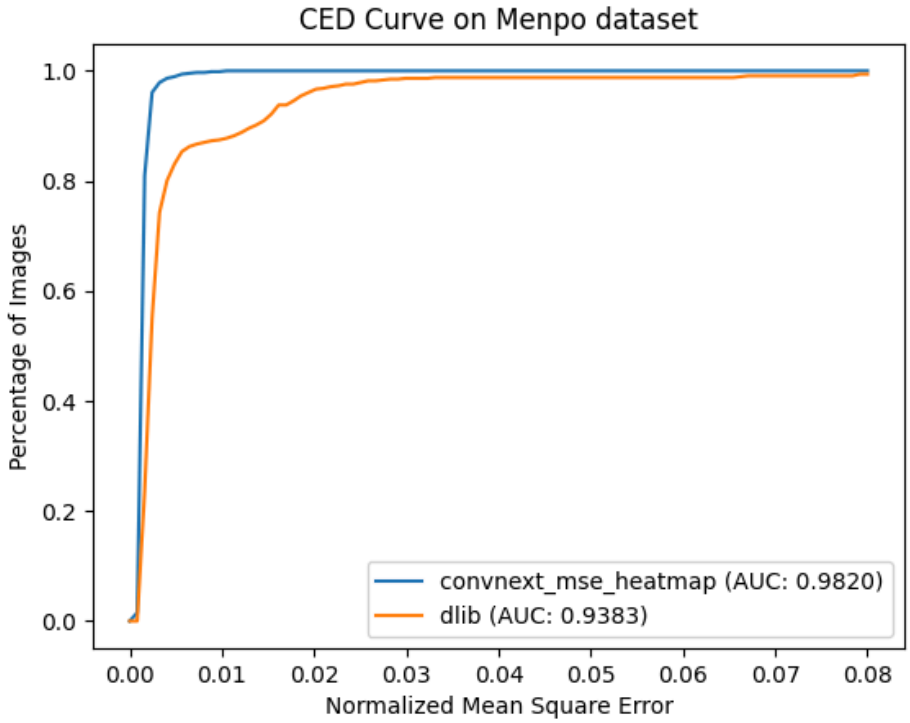
ConvNeXt + AdaptiveWing + Regression



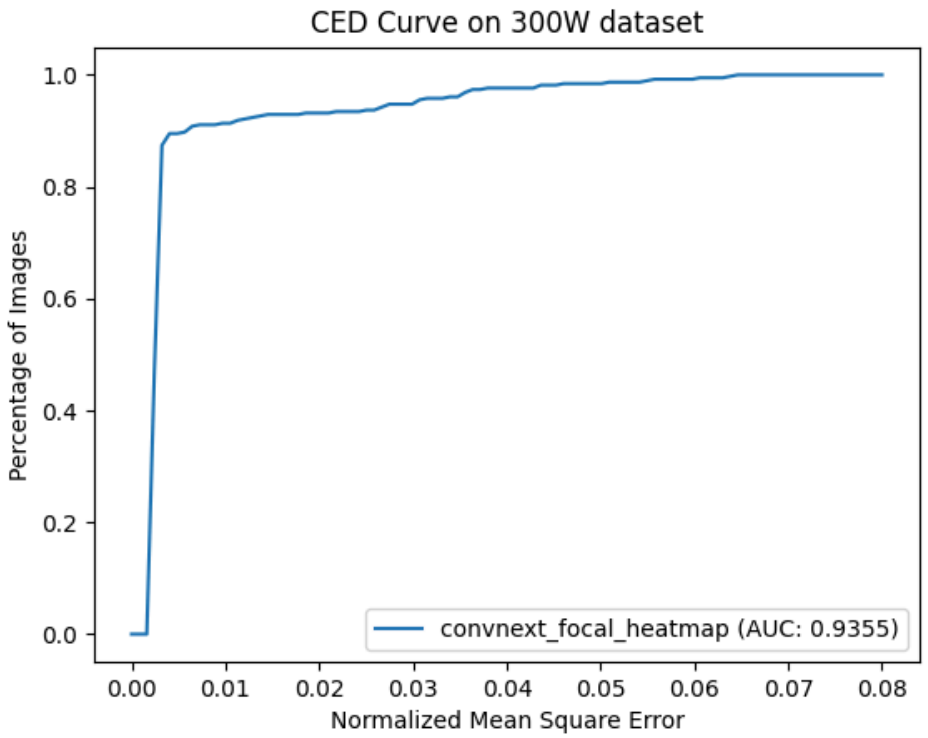


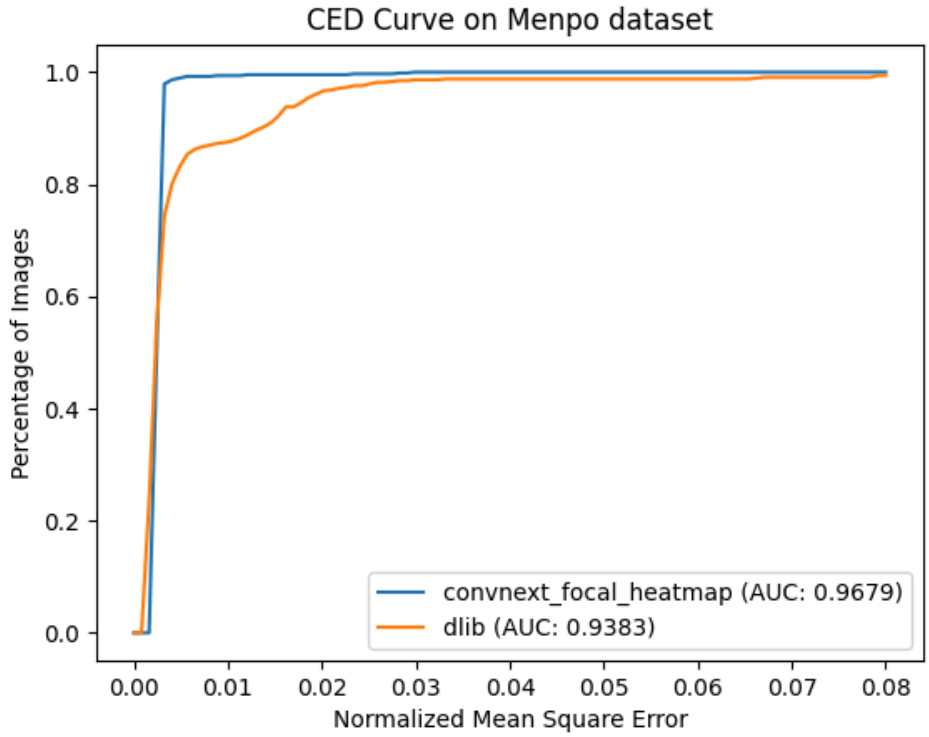
ConvNeXt + MSE + Heatmap





ConvNeXt + FocalLoss + Heatmap





ConvNeXt + BCE + Heatmap