Improving Standard Transformer Models for 3D Point Cloud Understanding with Image Pretraining

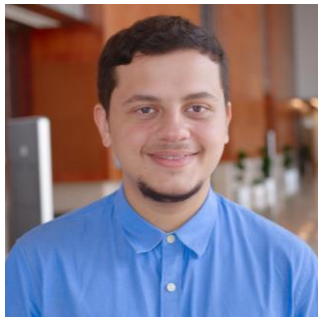
Рецензия

Федорова Анна, БПМИ-191

Авторы

Guocheng Qian

Xingdi Zhang



Abdullah Hamdi

Bernard Ghanem

Важные работы:

- PointNeXt (Guocheng Qian и Bernard Ghanem)
- Voint cloud (Abdullah Hamdi и Bernard Ghanem)

Некоторые модели в области

- PointNet (2016)
- PointNet++ (2017)
- Point-BERT (2021)
- PointMLP (2022, февраль)
- Point-MAE (2022, март)
- **PointNeXt** (2022, июнь)
- PViT, Pix4Point (2022, август)

SOTA

PointNet (2016)

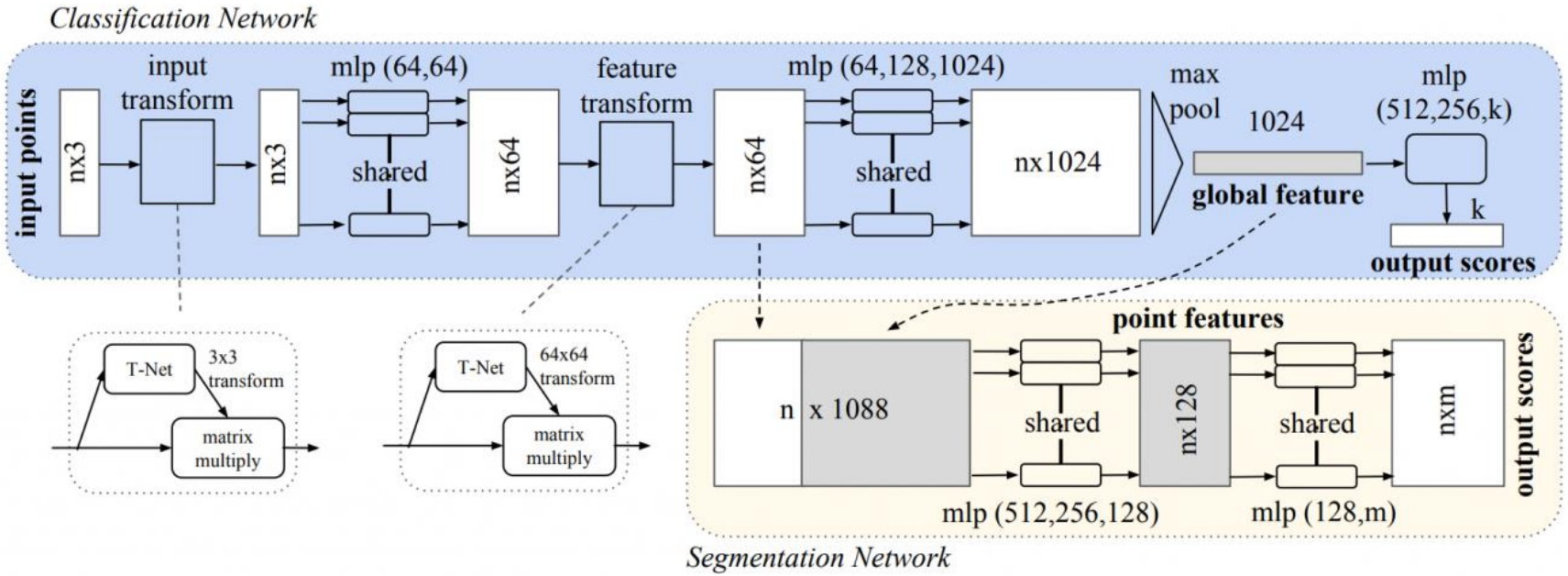


Figure 2. **PointNet Architecture.** The classification network takes n points as input, applies input and feature transformations, and then aggregates point features by max pooling. The output is classification scores for k classes. The segmentation network is an extension to the classification net. It concatenates global and local features and outputs per point scores. "mlp" stands for multi-layer perceptron, numbers in bracket are layer sizes. Batchnorm is used for all layers with ReLU. Dropout layers are used for the last mlp in classification net.

PointNet++ (2017)

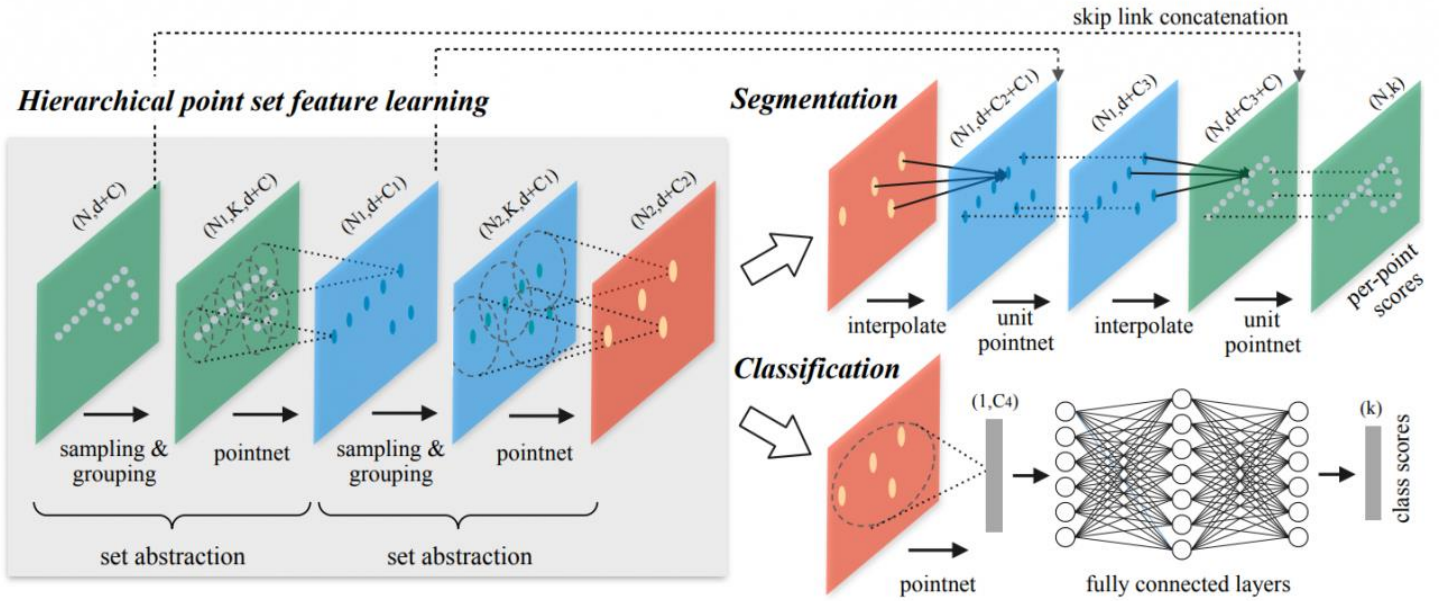


Figure 2: Illustration of our hierarchical feature learning architecture and its application for set segmentation and classification using points in 2D Euclidean space as an example. Single scale point grouping is visualized here. For details on density adaptive grouping, see Fig. 3

Улучшения PointNet++

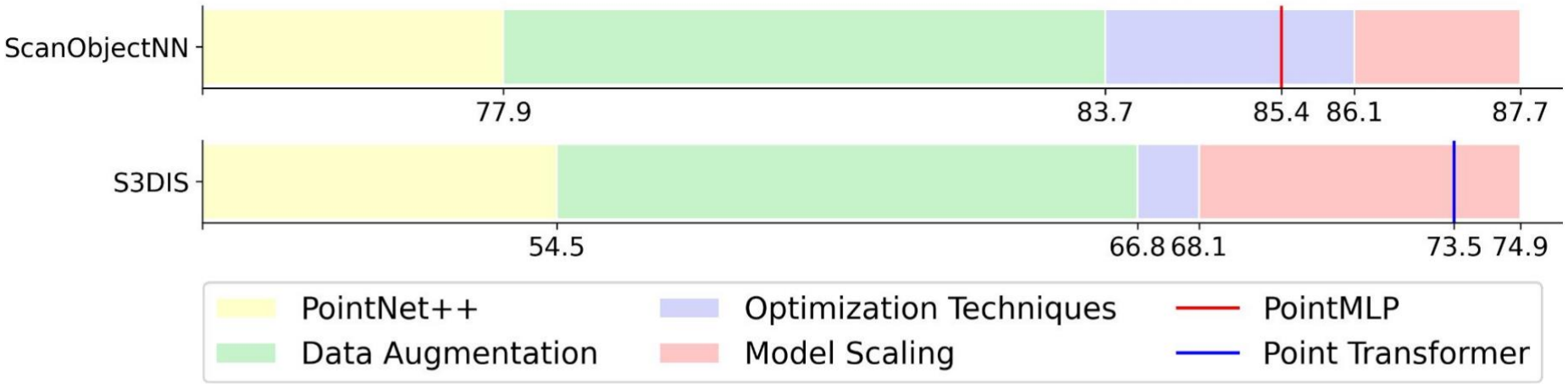


Figure 1: **Effects of training strategies and model scaling on PointNet++** [30]. We show that improved training strategies (data augmentation and optimization techniques) and model scaling can significantly boost PointNet++ performance. The average overall accuracy and mIoU (6-fold cross-validation) are reported on ScanObjectNN [44] and S3DIS [1].

PointNeXt – SOTA для 3D облаков точек

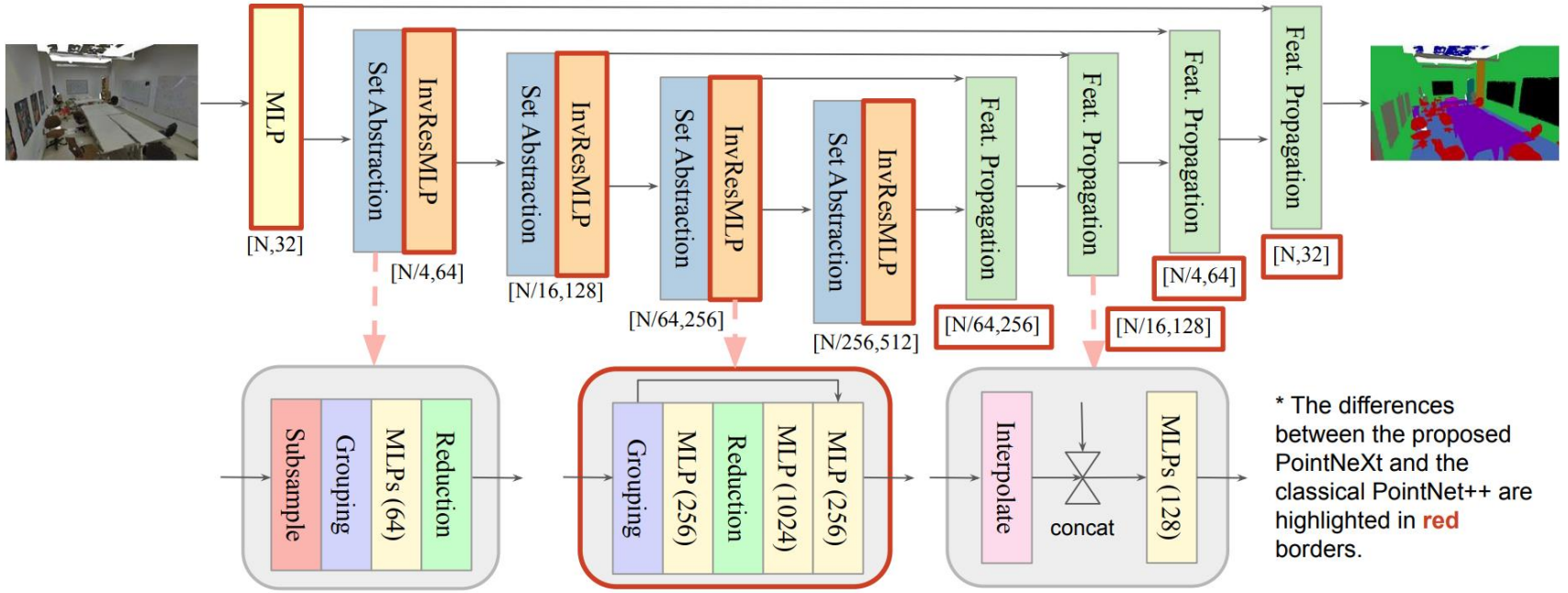
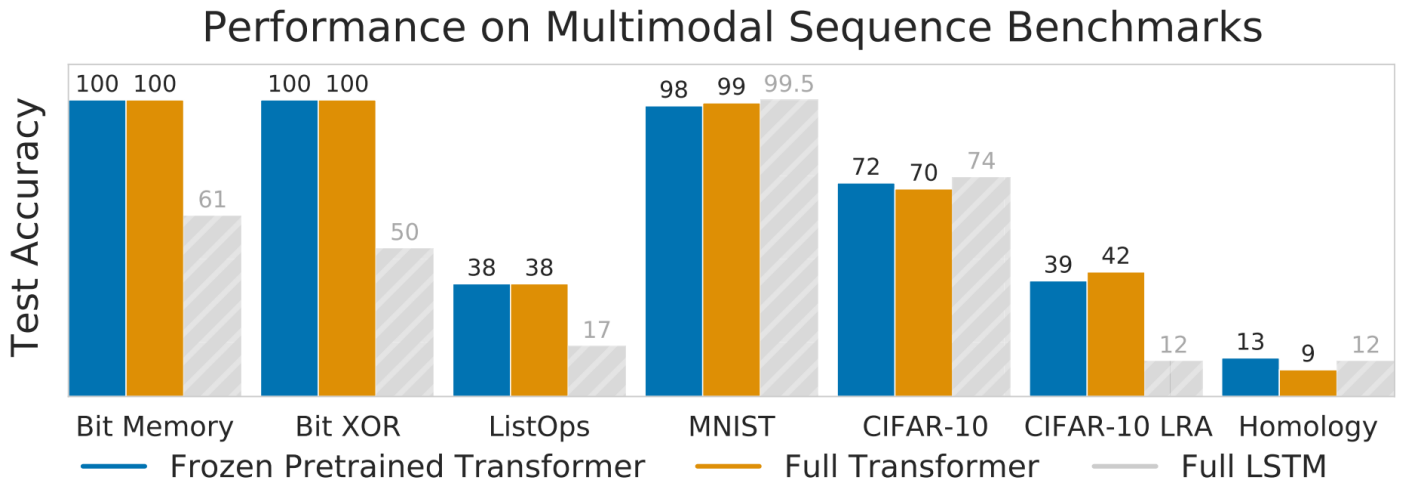


Figure 2: **PointNeXt architecture.** PointNeXt shares the same Set Abstraction and Feature Propagation blocks as PointNet++ [30], while adding an additional MLP layer at the beginning and scaling the architecture with the proposed Inverted Residual MLP (InvResMLP) blocks.

Pretrained Transformers as Universal Computation Engines (2021)



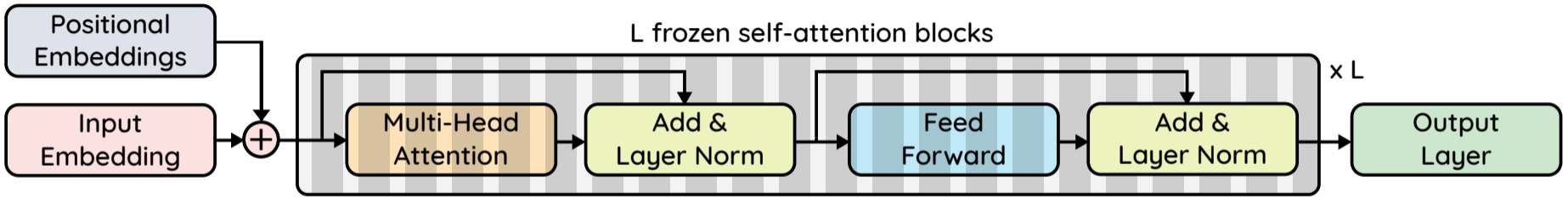


Figure 2: Frozen Pretrained Transformer (FPT). The self-attention & feedforward layers are frozen.

Итого

Плюсы
- Есть код
- Сравнивают на классических, но сложных задачах
- Близко к SOTA
- Трансформеры

Минусы
- Говорят, код не очень запускается
- Аргументируя transfer learning между областями ссылаются на спорную статью
- Не SOTA :(

Возможное продолжение
- Transfer learning из изображений в облака точек для других моделей
- Изменение стратегии обучения PViT

Всё!