

3) 실험 결과

- Fine-tuning on a Vision Task

- ✓ “Title” 카테고리에서 큰 성능 향상을 보임
 - ✓ “Title” 카테고리 : 다른 카테고리에 비해 영역이 작으며, textual content가 중요한 역할을 함
- Language Modality를 사전학습 과정에서 학습한 것이 성능 향상에 기여했을 것!

Model	Framework	Backbone	Text	Title	List	Table	Figure	Overall
PubLayNet[59]	Mask R-CNN	ResNet-101	91.6	84.0	88.6	96.0	94.9	91.0
DiT _{BASE} [30]	Mask R-CNN	Transformer	93.4	87.1	92.9	97.3	96.7	93.5
UDoc [14]	Faster R-CNN	ResNet-50	93.9	88.5	93.7	97.3	96.4	93.9
DiT _{BASE} [30]	Cascade R-CNN	Transformer	94.4	88.9	94.8	97.6	96.9	94.5
LayoutLMv3_{BASE} (Ours)	Cascade R-CNN	Transformer	94.5	90.6	95.5	97.9	97.0	95.1