

Do Better ImageNet Models Transfer Better? (2018)

Kornblith, Simon, Jonathon Shlens, and Quoc V. Le. "Do better imagenet models transfer better?." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. Google Brain

IIE8557-01 동적계획법과 강화학습

경영과학연구실 이태헌

layer features that are much worse for transfer learning [한글 번역](#)



우리는 이 관계가 ImageNet에서 네트워크가 어떻게 훈련되는지에 매우 민감하다는 것을 발견했습니다; 많은 일반적인 정규화 형태들이 ImageNet 정확도를 약간 향상시키지만, 전이 학습에 있어서는 마지막에서 두 번째 계층의 특성이 훨씬 더 나쁜 결과를 가져옵니다.



Additionally, we find that, on two small fine-grained image classification datasets, pretraining on ImageNet provides minimal benefits, indicating the learned features from ImageNet do not transfer well to fine-grained tasks. [한글 번역](#)



게다가, 우리는 두 개의 작은 세부 이미지 분류 데이터셋에서 ImageNet에 대한 사전 훈련이 최소한의 이점을 제공한다는 것을 발견했습니다, 이는 ImageNet에서 학습된 특성이 세부 작업으로 잘 전이되지 않음을 나타냅니다.



Together, our results show that ImageNet architectures generalize well across datasets, but ImageNet features are less general than previously suggested. [한글 번역](#)



결합하여, 우리의 결과들은 ImageNet 아키텍처가 데이터셋 간에 잘 일반화된다는 것을 보여주지만, ImageNet 특성은 이전에 제안된 것보다 덜 일반적임을 나타냅니다.



너가 방금 말한 대답 좀 더 쉽게 설명해줘



당연합니다. 이 결과는 ImageNet 모델(아키텍처)이 다양한 데이터셋에 잘 적용될 수 있다는 것을 보여주지만, ImageNet에서 배운 특성들이 다른 곳에 적용하기에는 그다지 적합하지 않을 수 있다는 것을 의미합니다. 다시 말해서, 모델의 구조는 여러 곳에서 유용하지만, ImageNet에서 배운 구체적인 정보나 패턴들은 다른 데이터셋이나 작업에선 그다지 유용하지 않을 수 있습니다.

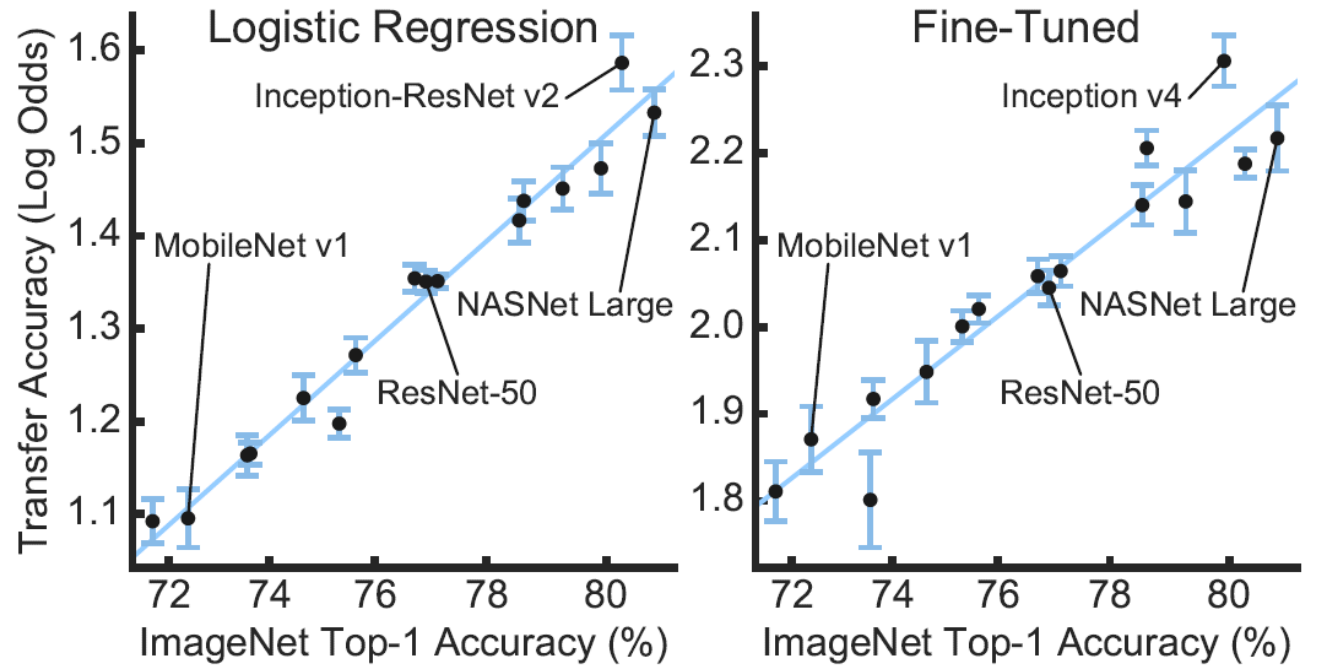
Regenerate

따라서, "Transfer Accuracy (Log Odds)" 값이 1.6일 경우, 해당 모델이 맞출 확률은 대략 83.2%입니다.

따라서, "Transfer Accuracy (Log Odds)" 값이 2.3일 경우, 해당 모델이 맞출 확률은 대략 90.9%입니다.

"Transfer Accuracy (Log Odds)"에서 "Log Odds"는 로짓(logit) 변환을 나타냅니다. 로짓 변환은 확률 값을 Odds 비로 변환한 뒤, 그 결과에 로그를 취한 값입니다.

전이 학습 성능은 고정된 ImageNet 특성(왼쪽)과 ImageNet 초기화에서의 미세 조정(오른쪽)에 대해 ImageNet top-1 정확도와 높은 상관관계를 보입니다. 각 플롯 내의 16개의 점은 12개의 데이터셋에 대해 로짓 변환 후 평균화된 16개의 다른 CNN 아키텍처에 대한 전이 정확도를 나타냅니다(섹션 3 참조)



ImageNet

- 컴퓨터 비전 연구에서 광범위하게 사용되는 대규모 이미지 데이터셋

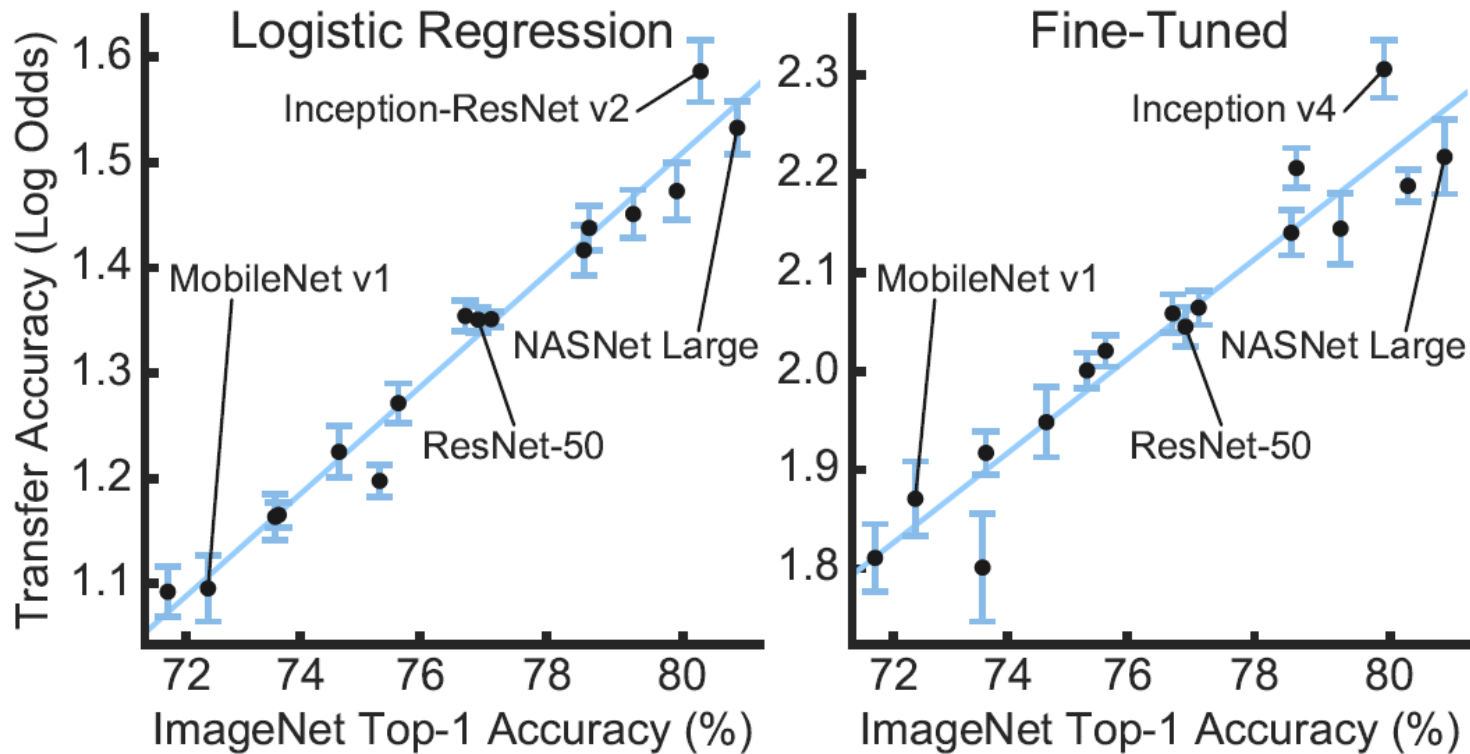


모델 아키텍처와 전이 관계 연구 필요성

- 현대 컴퓨터 비전 연구에서의 가설은 'ImageNet에서 더 좋은 성능을 내는 모델은 다른 비전 작업에서도 더 나은 성능을 보인다' 임
- 전이 학습은 컴퓨터 비전 분야의 기초이지만, 아키텍처와 전이 간의 관계를 평가하기 위한 연구는 많이 이루어지지 않음

모델 아키텍처와 전이 관계 연구 필요성

- 전이 학습 성능은 Fixed ImageNet feature (왼)
- ImageNet fine-tuning from ImageNet initialization (오)에 대해 ImageNet top-1 정확도와 높은 상관관계를 보임.



**‘ImageNet에서 더 좋은 성능을 내는 모델은
다른 비전 작업에서도 더 나은 성능을 보인다’는 가설을 테스트하고자 함**

실험 설계

- ImageNet 특성과 ImageNet 분류 아키텍처에 따라 실험 설계

전이 학습에 사용된 CNN classification model

◀ Inception v1	▼ Inception v4	– ResNet-50	∩ DenseNet-121	♣ MobileNet v1	■ NASNet-A Mobile
▲ BN-Inception	▽ Inception-ResNet v2	ResNet-101	∪ DenseNet-169	♠ MobileNet v2	★ NASNet-A Large
▶ Inception v3		× ResNet-152	∩ DenseNet-201	● MobileNet v2 (1.4x)	

전이 학습에 사용된 데이터셋

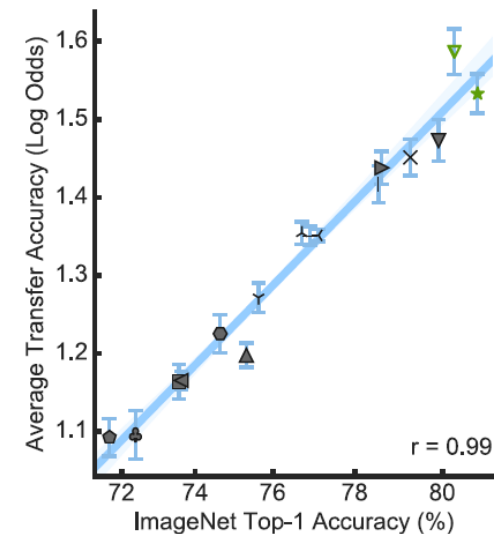
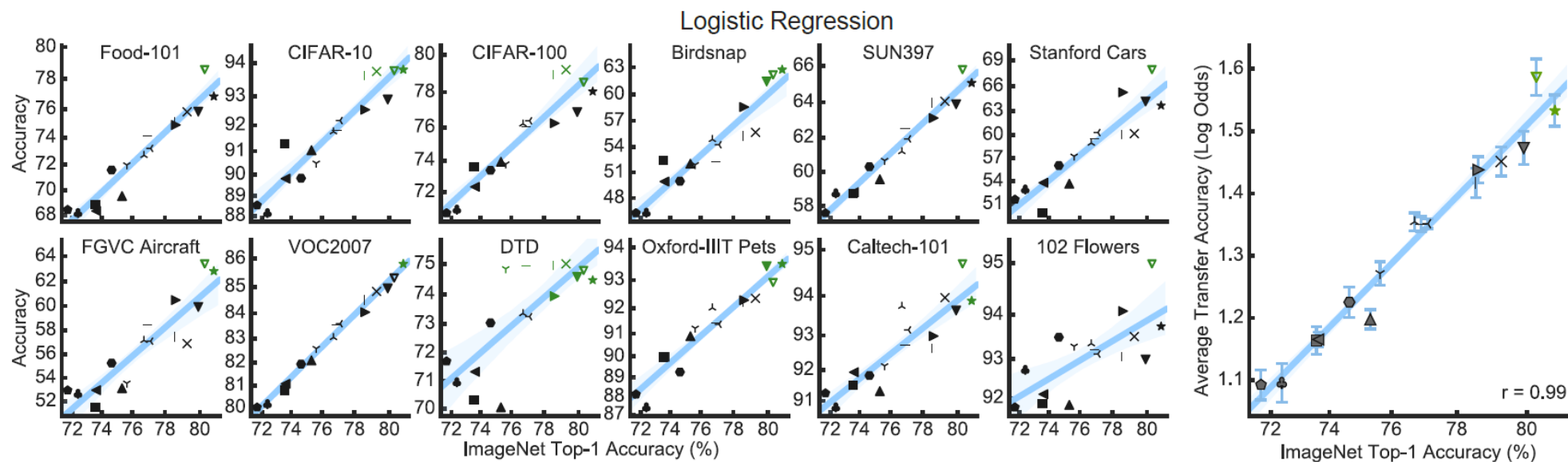
Dataset	Classes	Size (train/test)	Accuracy measure
Food-101 [5]	101	75,750/25,250	top-1
CIFAR-10 [37]	10	50,000/10,000	top-1
CIFAR-100 [37]	10	50,000/10,000	top-1
Birdsnap [4]	500	47,386/2,443	top-1
SUN397 [72]	397	19,850/19,850	top-1
Stanford Cars [36]	196	8,144/8,041	top-1
FGVC Aircraft [48]	100	6,667/3,333	mean per-class
PASCAL VOC 2007 Cls. [19]	20	5,011/4,952	11-point mAP
Describable Textures (DTD) [10]	47	3,760/1,880	top-1
Oxford-IIIT Pets [53]	37	3,680/3,369	mean per-class
Caltech-101 [20]	102	3,060/6,084	mean per-class
Oxford 102 Flowers [52]	102	2,040/6,149	mean per-class

실험 세팅

- Fixed feature extractors
- Fine-tuned from ImageNet initialization
- Trained from random initialization

ImageNet accuracy predicts performance of logistic regression on fixed features

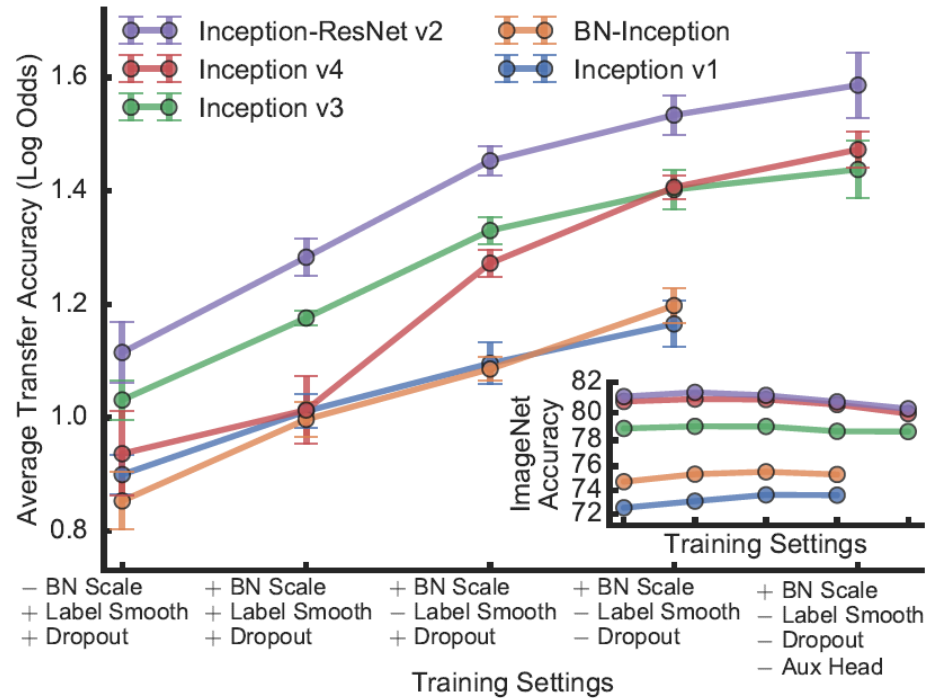
- ResNets and DenseNets consistently achieved higher accuracy than other models
- Correlation between ImageNet accuracy and transfer accuracy with fixed features was low and not statistically
- 낮은 상관관계는 공개된 checkpoint에 사용된 정규화 차이



Fixed feature extractor 에 정규화가 미치는 부정적 영향

- Inception 모델이 더 많은 정규화 기법을 사용하는 반면, ResNets, DenseNets은 상대적으로 덜 정규화 된 구성 사용

Transfer learning performance of Inception models with different training settings

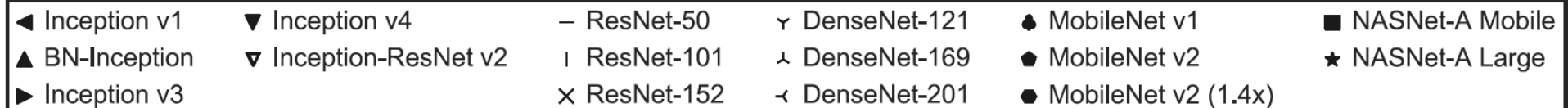
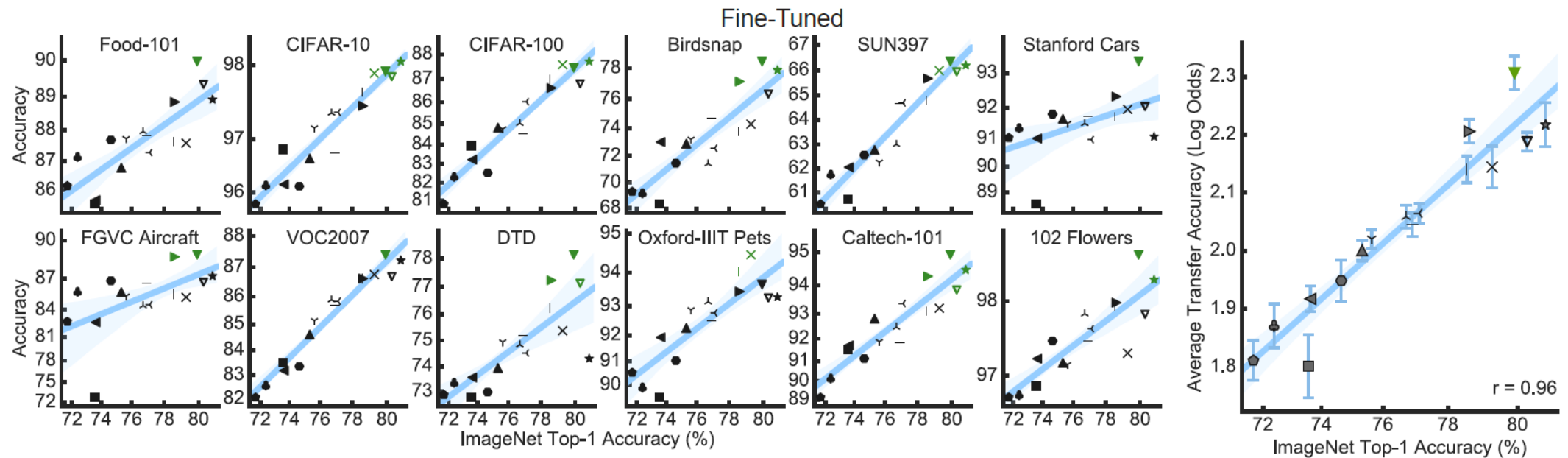


- 1) The absence of scale parameter (γ) for batch normalization layers
- 2) The use of label smoothing
- 3) The use of dropout
- 4) The presence of an auxiliary classifier head

- "+" indicates that a setting was enabled
- "-" indicates that a setting was disabled

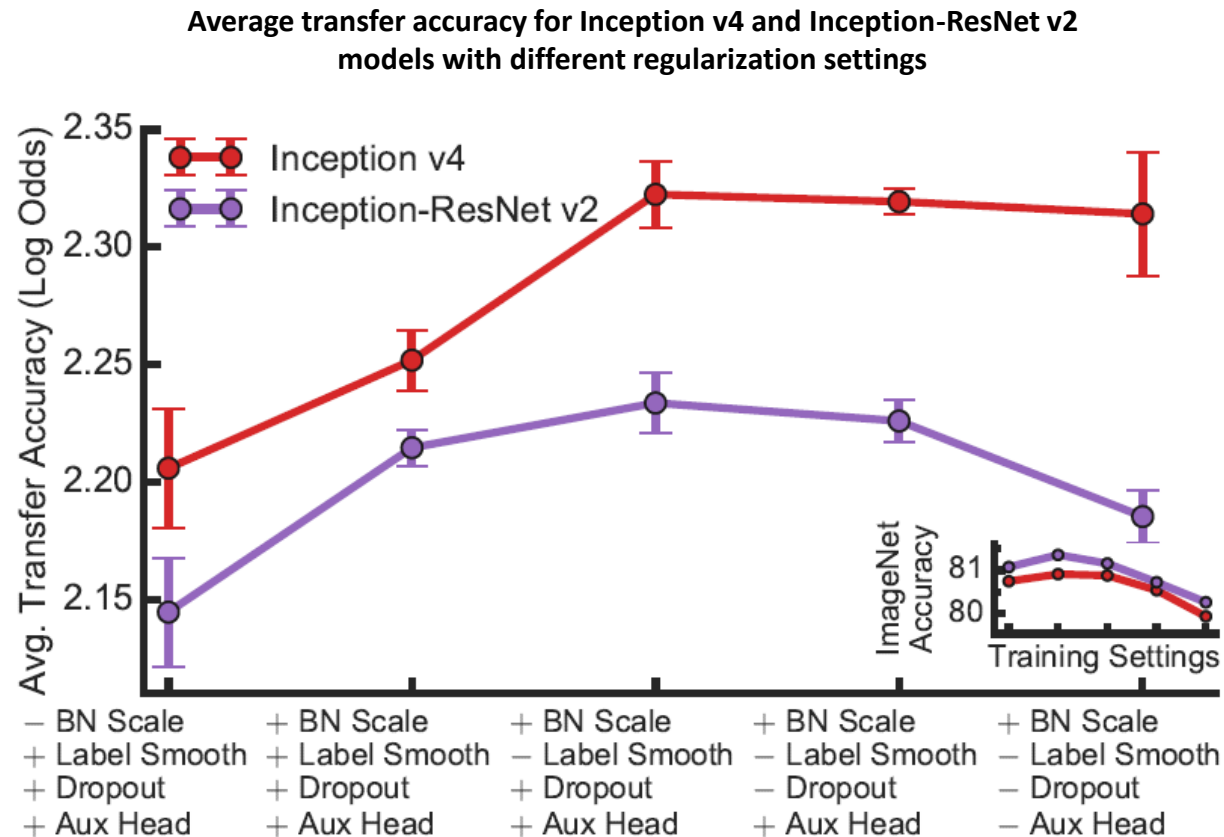
ImageNet accuracy predicts finetuning performance

- ImageNet top-1 accuracy was highly correlated with transfer accuracy



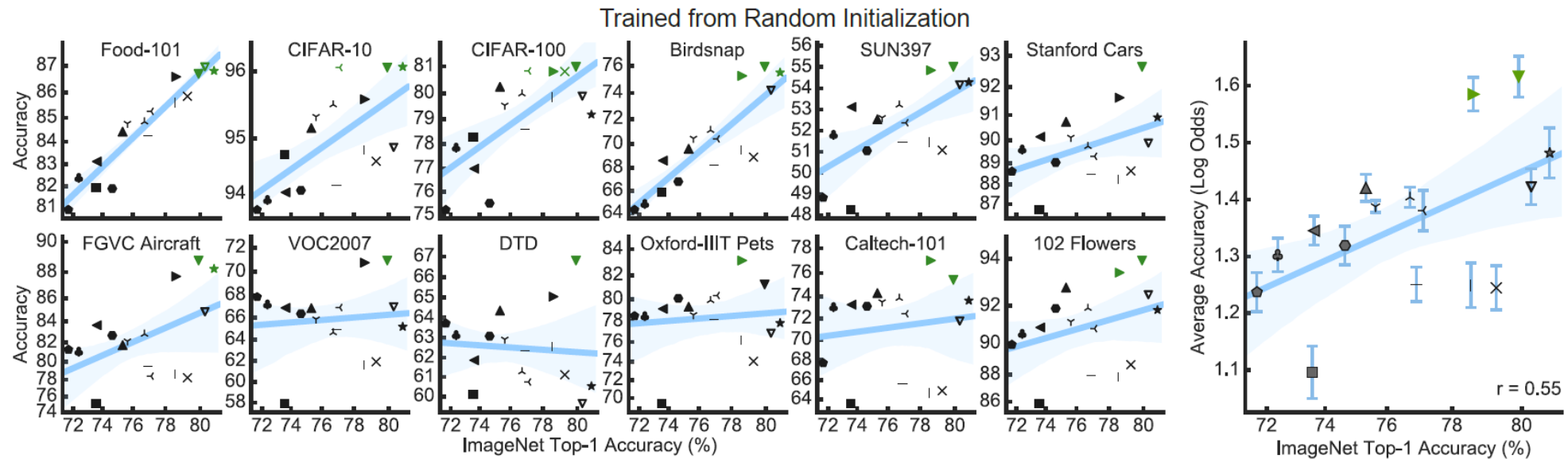
Fine-tuned models 에 정규화가 미치는 영향

- Fixed feature extractor models 와 비교하였을 때 정규화와 training settings는 fine-tuned models 성능에 적은 영향을 미침
- Fixed feature extractor models 실험과 마찬가지로, 배치 정규화 스케일 파라미터 도입하고 라벨 스무딩을 비활성화하면 성능이 향상됨



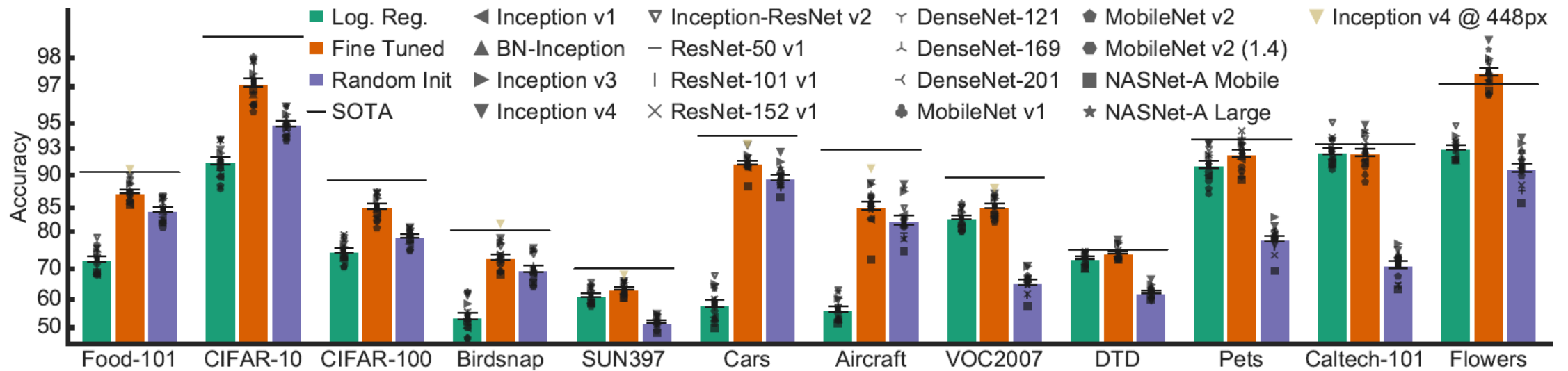
ImageNet accuracy predicts performance of networks trained from random initialization

- The correlation between ImageNet top-1 accuracy and accuracy on the new tasks was more variable than in the transfer learning settings
- But there was a tendency toward higher performance for models that achieved higher accuracy on ImageNet



Performance comparison of logistic regression, fine-tuning, and training from random initialization

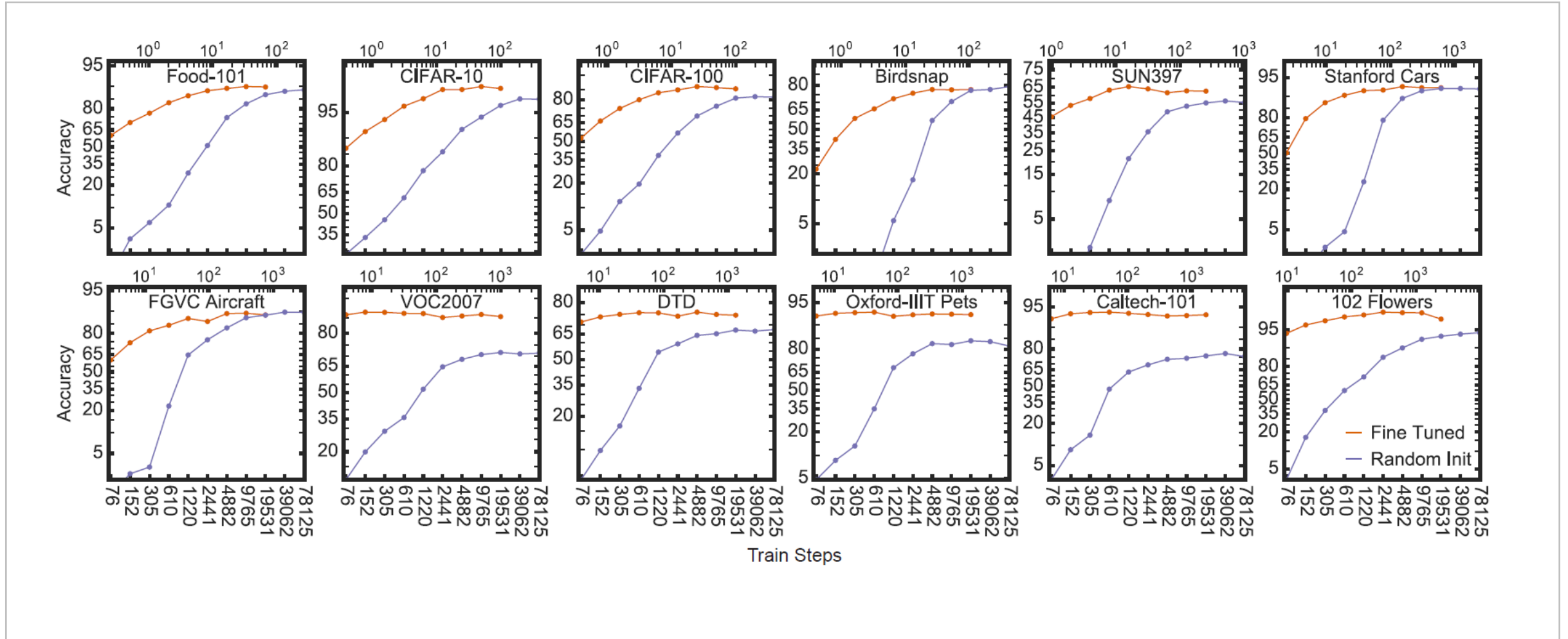
- Fine-tuning은 192개의 데이터셋, 모델 조합 중 179개에서 Fixed-feature 모델에 비해 성능을 향상 시킴
- Fine-tuning은 평균 성능에서 좋은 결과를 보임. 일반적으로 더 큰 데이터셋에서 더욱 성능이 향상됨



- Points represent individual models. Lines represent previous state-of-the-art

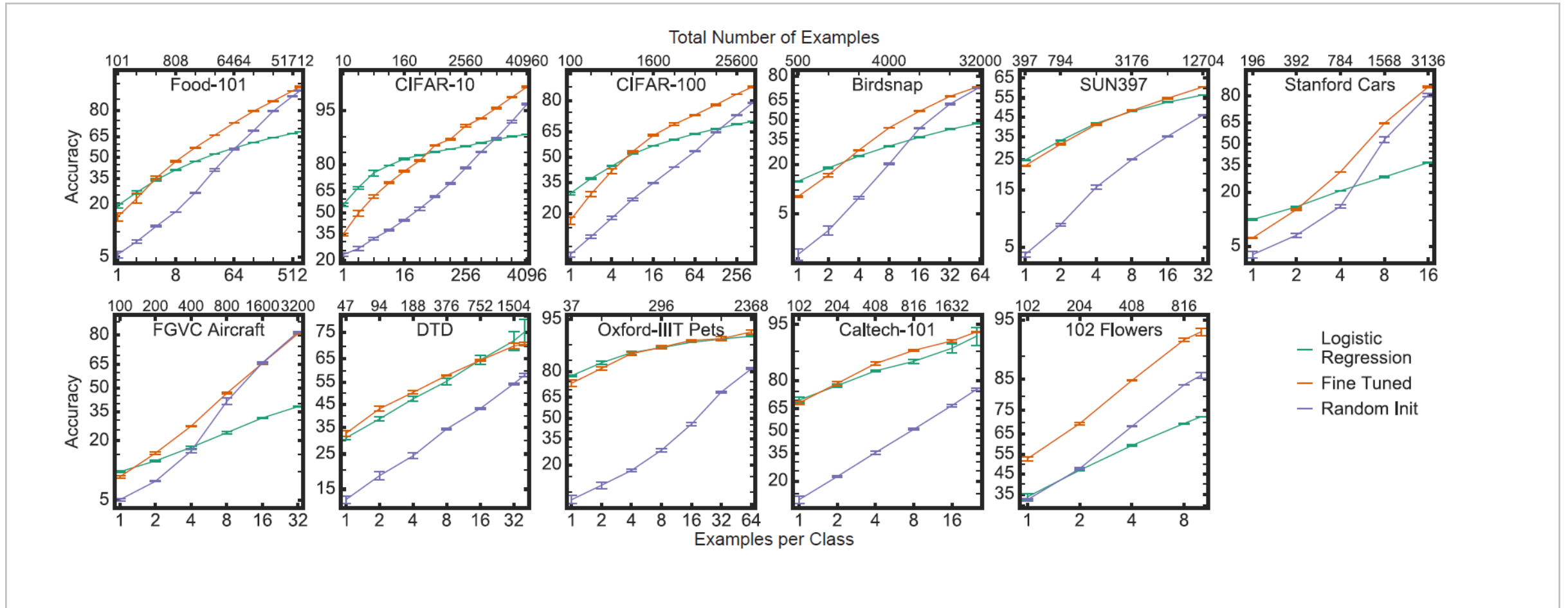
ImageNet pretraining accelerates convergence

- Performance of Inception v4 when fine-tuning or training from random initialization for different numbers of steps



Performance of transfer learning with the public Inception v4 model at different dataset sizes

- Pretraining on ImageNet improves performance on fine-grained tasks with small amounts of data, but the gap narrows quickly as dataset size increases



Conclusions

- 'ImageNet에서 더 좋은 성능을 내는 모델은 다른 비전 작업에서도 더 나은 성능을 보인다'는 가설을 검증함
- ImageNet에서의 좋은 성능과 transfer learning 정확도 사이에는 강한 상관관계가 있음
- Random initialization task에서도 ImageNet 데이터에 대한 성능이 좋은 모델이 Transfer 정확도가 높음. 이는 Computer vision 모델들이 ImageNet 데이터셋에 과적합 되어있지 않음을 확인할 수 있음
- 따라서, 성능이 좋은 ImageNet 아키텍처들이 더 나은 전이 가능한 표현들을 학습할 수 있음을 의미함

Q & A