

그린 AI

로이 슈워츠* ♦ 제시 닛지* ♦ ♣ 노아 A. 스미스 ♦ ♥ 오렌 에치오니♦

♦ Allen Institute for AI, 미국 워싱턴주 시애틀
♣ 카네기멜론대학교, 미국 펜실베이니아주 피츠버그
♥ 워싱턴 대학교, 시애틀, 워싱턴, 미국

2019년 7월

추상적인

딥 러닝 연구에 필요한 계산은 몇 달마다 두 배로 증가하여 2012년에서 2018년까지 약 300,000배 증가했습니다[2]. 이러한 계산에는 놀라운 정도로 큰 탄소 발자국이 있습니다[40]. 아이러니하게도 딥 러닝은 놀랍도록 에너지 효율적인 인간의 두뇌에서 영감을 받았습니다. 더욱이 계산의 재정적 비용으로 인해 학계, 학생 및 연구원, 특히 신흥 경제국의 연구자가 딥 러닝 연구에 참여하는 것이 어려울 수 있습니다.

본 성명서는 정확도 및 관련 조치와 함께 효율성을 연구의 평가 기준으로 삼아 실질적인 해결책을 제시합니다. 또한 점점 더 효율적인 방법을 조사하기 위한 기준선을 제공하기 위해 모델 개발, 교육 및 실행의 재정적 비용 또는 "가격표"를 보고할 것을 제안합니다. 우리의 목표는 AI를 보다 환경 친화적이고 포괄적으로 만드는 것입니다. 영감을 받은 학부생이라면 노트북을 사용하여 고품질 연구 논문을 작성할 수 있습니다. Green AI는 Allen Institute for AI의 새로운 초점입니다.

1 소개 및 동기

2012년 이후 인공지능 분야는 객체 인식, 게임 플레이, 기계 번역 등을 포함한 광범위한 기능에 대한 놀라운 발전을 보고했습니다[36]. 이러한 발전은 점점 더 커지고 계산 집약적인 딥 러닝 모델에 의해 달성되었습니다.1 [2]에서 재현한 그림 1은 2012년 AlexNet[20]에서 시작하여 최첨단 딥 러닝 모델에 대한 시간 경과에 따른 교육 비용 증가를 보여줍니다. 2017년에 AlphaZero [38]. 차트는 몇 달마다 교육 비용이 두 배로 증가하면서 전체적으로 300,000배의 증가를 보여줍니다. ELMo[29], BERT[8], openGPT-2[30], XLNet[48]을 살펴보면 NLP 단어 임베딩 접근 방식에서 훨씬 더 날카로운 경향을 관찰할 수 있습니다. 중요한 논문[40]은 여러 NLP 모델의 탄소 발자국을 추정했으며 이러한 추세가 환경 친화적이지 않고(Red AI 라고 함) 비용이 많이 들기 때문에 NLP 연구 참여에 대한 장벽을 높인다고 주장했습니다.

이러한 추세는 AI 커뮤니티가 "최첨단" 결과를 얻는 데 중점을 두기 때문에 발생합니다. 2 일반적으로 정확도 측정을 보고하지만 그에 대한 언급은 생략하는 리더보드[46, 45]의 인기가 높아지는 것이 그 예입니다. 비용 또는 효율성(예: leaderboards.allenai.org 참조). AI에서 모델 정확도 개선의 분명한 이점에도 불구하고 이 단일 지표에 초점을 맞추는 것은 보고된 정확도에 도달하는 경제적, 환경적 또는 사회적 비용을 무시합니다.

우리는 보다 환경 친화적이고 포괄적인 AI 연구인 Green AI에서 연구 활동을 늘리는 것을 지지합니다. 우리는 Red AI 연구가 AI 분야에 귀중한 기여를 해왔지만 지나치게 지배적이라는 점을 강조합니다. 우리는 녹색 AI 옵션으로 균형을 이동하여 영감을 받은 학부생이 노트북을 사용하여 최고의 연구 컨퍼런스에서 승인될 수 있는 고품질 논문을 작성할 수 있는 기회를 갖도록 하고자 합니다.

처음 두 저자가 동등하게 기여했습니다. 이 연구는 Allen Institute for AI에서 수행되었습니다.

1 간결함을 위해 이 백서 전체에서 AI를 언급하지만 딥 러닝 방법에 의존하는 AI 연구에 초점을 맞춥니다.

2 실제로 일부 벤치마크에서 시스템의 정확도가 이전에 보고된 시스템의 정확도보다 높다는 것을 의미합니다.

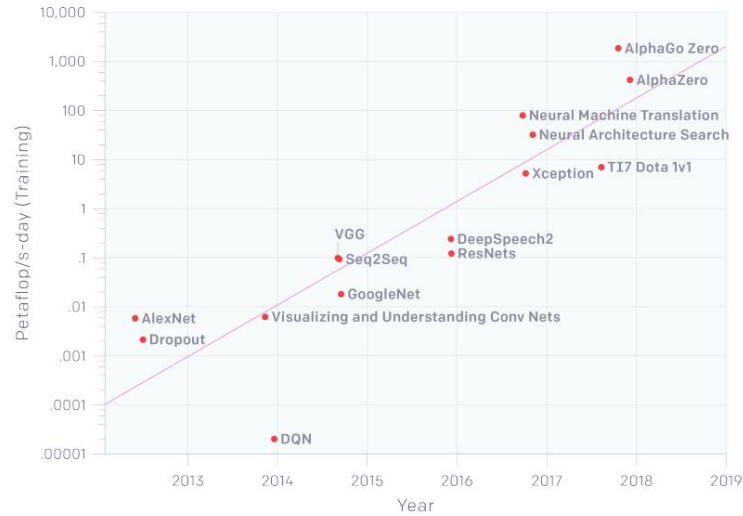


그림 1: 딥 러닝 모델 훈련에 사용되는 컴퓨팅 양이 6년 동안 300,000배 증가했습니다. [2]에서 가져온 그림.

구체적으로, 정확도 및 관련 측정과 함께 효율성을 AI 논문에 대한 보다 일반적인 평가 기준으로 만들 것을 제안합니다.

AI 연구는 여러 가지 방법으로 계산 비용이 많이 들 수 있지만 각각 효율적인 개선을 위한 기회를 제공합니다. 예를 들어, 계산 비용과 트레이닝 세트 크기의 함수로 정확성을 플롯하기 위해 논문이 필요할 수 있으며, 이는 미래에 보다 효율적인 연구를 위한 기준선을 제공합니다. 모델 찾기, 교육 및 실행에 대한 계산 비용 태그를 보고하는 것은 **Green AI**의 핵심 관행입니다(공식 1 참조). 투명성을 제공하는 것 외에도 가격표는 다른 연구자들이 개선할 수 있는 기준선입니다.

그림 2의 경험적 분석은 AI 연구 커뮤니티가 계산 효율성에 상대적으로 거의 관심을 기울이지 않았음을 시사합니다. 실제로 그림 1에서 알 수 있듯이 연구의 계산 비용은 무어의 법칙을 훨씬 능가하는 속도로 기하급수적으로 증가하고 있습니다[28]. **Red AI**는 잘 알려진 비용 증가로 인한 수익 감소에도 불구하고 증가하고 있습니다(예: 그림 3). 이 백서에서는 **Red AI**에 기여하는 핵심 요소를 식별하고 일부 AI 연구를 더 친환경적이고 포괄적이며 아마도 인지적으로 더 그럴듯하게 만드는 데 도움이 될 수 있는 간단하고 계산하기 쉬운 효율성 메트릭의 도입을 옹호합니다. **Green AI**는 환경 친화적인 과학 연구에 대한 광범위하고 장기적인 관심의 일부입니다(예: Green Chemistry 저널 참조). 특히 컴퓨터 과학은 지속 가능하고 에너지 효율적인 컴퓨팅을 조사한 오랜 역사를 가지고 있습니다(예: 지속 가능한 컴퓨팅: 정보학 및 시스템 저널 참조).

이 논문의 나머지 부분은 다음과 같이 구성된다. 섹션 2에서는 딥 러닝 연구를 **Red AI** 영역으로 옮기는 사례를 분석합니다. 섹션 3에서는 **Green AI**에 대한 우리의 제안을 논의합니다. 4절에서는 관련 연구를 살펴보고 향후 연구 방향에 대한 논의로 결론을 맺는다.

2 레드 AI

Red AI는 본질적으로 더 강력한 결과를 "구매"하는 대규모 계산 능력을 사용하여 정확도(또는 관련 조치)에서 최첨단 결과를 얻으려는 AI 연구를 말합니다. 그러나 모델 성능과 모델 복잡성(매개변수의 수 또는 추론 시간으로 측정됨) 사이의 관계는 기껏해야 대수적인 것으로 오랫동안 이해되어 왔습니다. 성능의 선형 이득을 위해서는 기하급수적으로 더 큰 모델이 필요합니다[18]. 훈련 데이터의 양[41, 13]과 실험 횟수[9]가 증가함에 따라 유사한 경향이 존재합니다. 이러한 각각의 경우에 수확 체감은 계산 비용을 증가시킵니다.

이 섹션에서는 **Red AI**에 기여하는 요인을 분석하고 시간이 지남에 따라 수익이 감소하는 결과를 보여줍니다(그림 3 참조). 우리는 **Red AI** 작업이 가치가 있으며 실제로 많은 부분이 우리가 알고 있는 것에 기여한다는 점을 다시 한 번 언급합니다.

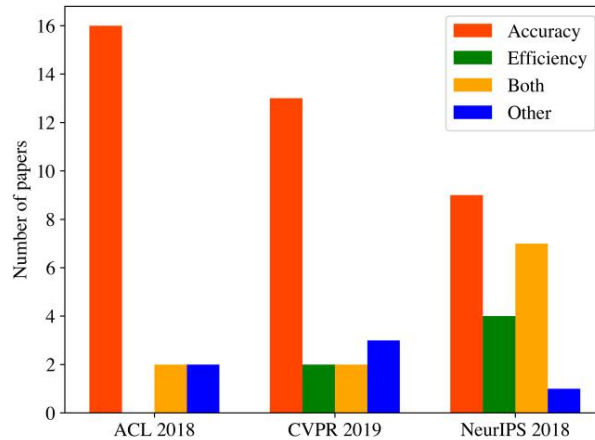


그림 2: AI 논문은 효율성보다는 정확도를 목표로 하는 경향이 있습니다. 이 수치는 상위 AI 컨퍼런스의 60개 논문 샘플에서 정확성, 효율성, 둘 다를 목표로 하는 논문의 비율을 보여줍니다.

AI의 한계를 뛰어넘다. 여기서 우리의 설명은 계산 비용이 높은 영역을 강조하고 각각을 보다 효율적인 기술을 개발할 수 있는 기회로 제시하기 위한 것입니다.

Red AI의 확산을 입증하기 위해 최고의 AI 컨퍼런스(ACL, 3 NeurIPS, 4 및 CVPR5)에서 60개의 논문을 샘플링했습니다. 각 논문에 대해 저자가 주장하는 주요 기여가 (a) 정확성 또는 일부 관련 조치의 개선인지, (b) 효율성의 개선인지, (c) 둘 다인지 또는 (d) 기타인지에 주목했습니다. 그림 2에서 볼 수 있듯이, 우리가 고려한 모든 회의에서 대부분의 논문은 정확도를 목표로 했습니다(ACL 논문의 90%, NeurIPS 논문의 80%, CVPR 논문의 75%). 더욱이 경험적 AI 컨퍼런스(ACL 및 CVPR)의 경우 단지 작은 부분(각각 10% 및 20%)만이 새로운 효율성 결과를 주장합니다.⁶ 이는 정확도와 같은 성능 측정에 대한 AI 커뮤니티의 초점을 강조합니다. 속도 또는 모델 크기와 같은 효율성 측정 비용. 본 논문에서는 후자에 더 큰 비중을 두어야 한다고 주장한다.

AI 연구가 **위험** 할 수 있는 다양한 방식을 더 잘 이해하려면 과학 논문에 보고된 AI 결과를 고려하십시오. 이 결과에는 일반적으로 훈련 데이터 세트에서 훈련되고 테스트 데이터 세트에서 평가된 모델이 포함됩니다. 해당 모델을 개발하는 과정에는 하이퍼파라미터를 조정하기 위한 여러 실험이 포함되는 경우가 많습니다. 이러한 결과를 생성하는 데 드는 계산 및 환경 비용을 증가시키는 다양한 요소를 고려할 때 세 가지 요소가 떠오릅니다. 훈련 중 모델이 실행되는 횟수를 제어하는 훈련 (D)ataset의 크기와 모델 개발 중에 모델이 훈련되는 횟수를 제어하는 (H)yperparameter 실험의 수입니다. 기계 학습에서 (R)결과를 생성하는 총 비용은 이러한 각 수량에 따라 선형적으로 증가합니다. 이 비용은 다음과 같이 추정할 수 있습니다.

$$\text{비용}(R) \propto E \cdot D \cdot H$$

방정식 1: **Red AI**의 방정식: AI(R)결과 비용은 단일 (E)예제 처리 비용, 훈련 (D)데이터 세트의 크기 및 (H)yperparameter 실험의 수에 따라 선형적으로 증가합니다.

방정식 1은 단순화된 것입니다(예: 다른 하이퍼파라미터 할당은 단일 예제를 처리하는 데 다른 비용으로 이어질 수 있음). 또한 교육 에폭 수와 같은 다른 요인도 무시합니다. 그럼에도 불구하고 세 가지를 보여줍니다.

³<https://acl2018.org>

⁴<https://nips.cc/Conferences/2018> ⁵<http://cvpr2019.thecvf.com>

⁶ 흥미롭게도 많은 NeurIPS 논문에는 성능을 예제 또는 반복의 함수로 설명하는 수렴률 또는 후회 범위가 포함되어 있어 효율성(55%)을 목표로 합니다. 이것은 적어도 이론적 분석에서 이 개념의 중요성에 대한 인식이 증가했음을 나타냅니다.

결과 생성의 총 비용에서 각각 중요한 요소인 양. 아래에서는 각 수량을 개별적으로 고려합니다.

하나의 예에 대한 비용이 많이 드는 처리 우리의 초점은 각 훈련 단계에서 추론이 필요한 것이 일반적인 신경 모델에 있으므로 예를 "처리"하는 것으로 훈련 및 추론 비용을 함께 논의합니다. 일부 작업에서는 많은 양의 리소스가 필요한 점점 더 비싼 모델을 사용했으며, 그 결과 이러한 모델에서 추론을 수행하려면 많은 계산과 훈련이 필요할 수 있습니다. 예를 들어 Google의 BERT-large[8]에는 약 3억 5천만 개의 매개변수가 포함되어 있습니다. openAI의 openGPT2-XL 모델[30]에는 15억 개의 매개변수가 포함되어 있습니다.

우리의 홈 조직인 AI2는 최근에 15억 개의 매개변수를 포함하는 Grover [49]를 출시했습니다. 컴퓨터 비전 커뮤니티에서도 유사한 추세가 관찰됩니다(그림 1).

이러한 대형 모델은 각 예제를 처리하는 데 드는 비용이 높기 때문에 교육 비용이 많이 듭니다. BERT-large는 4일 동안 64개의 TPU 칩에서 학습되었습니다. Grover는 256개의 TPU 칩에 대해 2주 동안 25,000달러의 예상 비용으로 교육을 받았습니다. XLNet은 BERT-large와 유사한 아키텍처를 가지고 있지만 더 비싼 목적 함수(10배 더 많은 데이터에 추가)를 사용했으며 2.5일 동안 512개의 TPU 칩에서 학습되었습니다.7 최고의 BERT-large를 재현하는 것은 불가능합니다. 단일 GPU를 사용하는 큰 결과8 또는 XLNet 결과9. 전문화된 모델은 시간당 \$1,000 이상의 비용으로 Go[37]의 단일 게임을 플레이하는 데 1,920개의 CPU와 280개의 GPU가 필요한 최고 버전인 AlphaGo와 같이 훨씬 더 극단적인 비용을 가질 수 있습니다.10

단일 모델의 변형(예: BERT-small 및 BERT-large)을 검토할 때 더 큰 모델이 더 강력한 성능을 가질 수 있음을 확인했으며 이는 귀중한 과학적 기여입니다. 그러나 이는 모델 성장 속도가 모델 성능의 증가를 훨씬 초과하기 때문에 점점 더 커지는 AI 모델의 재정적 및 환경적 비용이 곧 감소하지 않을 것임을 의미합니다[16]. 결과적으로 단순히 AI 모델을 더 크게 만들어 AI 모델을 계속 개선하려면 점점 더 많은 리소스가 필요할 것입니다.

많은 예제 처리 최근 AI에서 최첨단 성능이 발전하고 있는 또 다른 방법은 훈련 데이터 모델의 양을 지속적으로 늘리는 것입니다. BERT-large는 30억 개의 단어 조각에 대한 교육 후 2018년에 많은 NLP 작업에서 최고의 성능을 보였습니다. XLNet은 최근 Common Crawl의 일부를 포함하여 320억 개의 단어 조각에 대한 학습 후 BERT를 능가했습니다. 400억 단어로 훈련된 openGPT-2-XL; FAIR의 RoBERTa[23]는 160GB의 텍스트, 약 400억 개의 단어 조각으로 훈련되었으며, 훈련하는 데 약 25,000 GPU 시간이 필요했습니다. 컴퓨터 비전에서 Facebook[25]의 연구원들은 Instagram의 35억 개 이미지에 대한 이미지 분류 모델을 사전 훈련했으며, 이는 Open Images와 같은 기존의 레이블이 지정된 이미지 데이터 세트보다 3배 더 큼니다. 이러한 모델의 결과 또는 동일한 설정에서 자체 모델을 교육합니다(특히 여러 시대에 대한 교육이 표준이므로). 예를 들어 2019년 6월 Common Crawl에는 242TB의 압축되지 않은 데이터가 포함되어12 데이터를 저장하는 것조차 비용이 많이 듭니다. 마지막으로, 모델 크기의 경우와 마찬가지로 성능 향상을 위해 더 많은 데이터에 의존하는 것은 더 많은 데이터를 추가하는 결과가 감소하기 때문에 비용이 많이 드는 것으로 악명이 높습니다[41]. 예를 들어, [25]에서 가져온 그림 3은 물체 인식 상위 1위 정확도와 훈련 예제 수 사이의 대수 관계를 보여줍니다.

엄청난 수의 실험 일부 프로젝트에서는 대부분의 연구자가 도달할 수 없는 초매개변수를 조정하거나 신경 구조를 검색하는 데 많은 양의 계산을 쏟아 부었습니다. 예를 들어, Google[51]의 연구원들은 신경 구조 검색에서 12,800개 이상의 신경망을 훈련하여 물체 감지 및 언어 모델링의 성능을 향상시켰습니다. 고정 아키텍처를 사용하여 DeepMind[26]의 연구자들은 1,500개의 하이퍼파라미터 할당을 평가하여 LSTM 언어 모델[15]이 최신 perplexity 결과에 도달할 수 있음을 입증했습니다. LSTM의 성능이 단지 몇 번의 하이퍼파라미터 시도 후에 정체되지 않는다는 것을 보여주는 이 결과의 가치에도 불구하고, 공정한 비교를 위해 다른 경쟁 모델의 잠재력을 완전히 탐색하는 것은 엄청나게 비용이 많이 듭니다.

7이 프로세스의 비용에 대한 일부 추정치는 \$250,000에 달합니다 (twitter.com/eturner303/status/1143174828804857856).
8<https://github.com/google-research/bert> 참조 9<https://github.com/zihangdai/xlnet> 참조 10최신 버전의 AlphaGo는 훨씬 더 효율적입니다[39]. 11<https://opensource.google.com/projects/open-images-dataset> 12<http://commoncrawl.org/2019/07/>

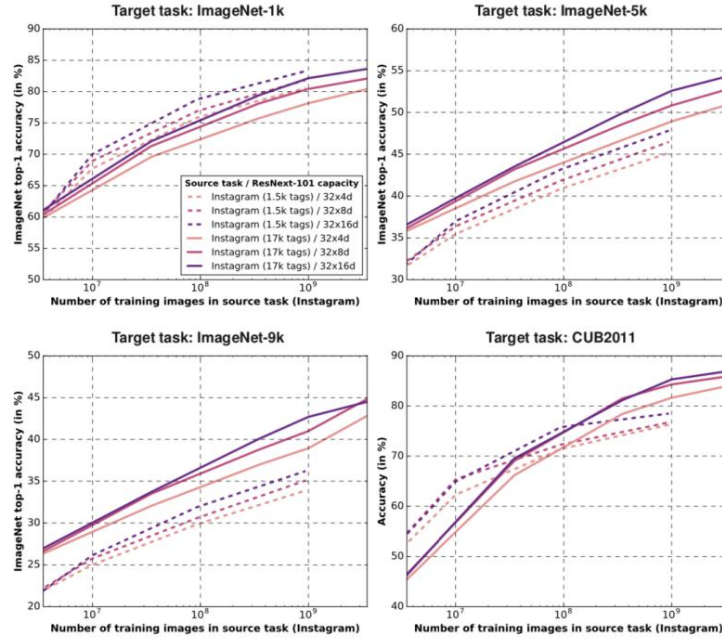


그림 3: 더 많은 데이터에 대한 교육의 수익 감소: 개체 감지 정확도는 교육 예제의 수가 기하급수적으로 증가함에 따라 선형적으로 증가합니다 [25].

엄청난 수의 실험이라는 주제는 위에서 논의한 처음 두 가지만큼 잘 연구되지 않았습니다. 사실, 모델 구성 중에 수행된 실험의 수는 종종 과소 보고됩니다. 그림에도 불구하고 여기에서도 실험 횟수와 성능 향상 사이에 로그 관계에 대한 증거가 존재합니다[9].

토론 모델에 더 많은 리소스를 투입함으로써 얻을 수 있는 이점은 확실히 AI 커뮤니티의 관심 대상입니다. 실제로 모델 크기, 데이터 세트 크기 및 하이퍼파라미터 검색 공간의 한계를 뛰어넘는 데 가치가 있습니다. 현재 최신 AI 모델에 막대한 양의 리소스를 투입했음에도 불구하고 이러한 투자는 다운스트림 성능 측면에서 여전히 성과를 거두고 있습니다(비록 점점 더 낮아지고 있지만). 포화점(있는 경우)을 찾는 것은 AI의 미래에 중요한 질문입니다.

이 백서의 목표는 Red AI의 비용에 대한 인식을 높이고 AI 커뮤니티가 정확성보다는 효율성을 최적화하여 다른 경로를 택하는 연구원의 작업 가치를 인식하도록 권장하는 것입니다. 다음으로 AI를 보다 친환경적으로 만들기 위한 구체적인 조치에 대해 논의합니다.

3 그린 AI

그린 AI라는 용어는 계산 비용을 늘리지 않고 이상적으로는 줄이지 않고 참신한 결과를 내는 AI 연구를 의미합니다. Red AI는 컴퓨팅(따라서 탄소) 비용을 빠르게 증가시키는 결과를 낳았지만 Green AI는 그 반대의 효과를 냅니다. 효율성 측정이 정확도와 함께 연구의 중요한 평가 지표로 널리 받아들여지면 연구자는 환경과 포용성 모두에 긍정적인 영향을 미치는 모델의 효율성에 집중할 수 있는 선택권을 갖게 됩니다. 이 섹션에서는 보고 및 최적화할 수 있는 효율성의 여러 척도를 검토하고 AI 연구 결과가 발표될 때 보고되어야 한다고 주장하는 특정 척도인 FPO를 지지합니다.

3.1 효율성 측정

효율성을 측정하기 위해 AI에서 결과를 생성하는 데 필요한 작업량, 즉 모델을 교육하는 데 필요한 작업량과 해당하는 경우 모든 하이퍼파라미터 튜닝 실험에 대한 작업의 합계를 보고하는 것이 좋습니다. 처럼

실험 비용은 단일 예를 처리하는 비용, 데이터 세트의 크기 및 실험 수(공식 1)로 분해되며, 이러한 각 단계에서 작업량을 줄이면 보다 친환경적인 시가 됩니다. .

모델이 수행한 작업량을 보고할 때 서로 다른 모델 간에 공정한 비교가 가능한 양을 측정하려고 합니다. 결과적으로 이 측정은 서로 다른 실험실에서 서로 다른 시간에 서로 다른 하드웨어를 사용하여 이상적으로 안정적이어야 합니다.

탄소배출량 탄소배출량은 우리가 직접적으로 최소화하고 싶은 양이라는 점에서 매력적입니다. 그럼에도 불구하고 모델을 훈련하거나 실행하여 방출되는 탄소의 정확한 양을 측정하고 그에 따라 AI 결과를 생성하는 것은 비실용적입니다. 이 양은 지역 전기 인프라에 크게 의존하기 때문입니다. 그 결과, 서로 다른 위치에 있는 연구원들 간에 또는 심지어 다른 시간에 같은 위치에 있는 연구원들 간에도 비교할 수 없습니다.

전기 사용량 전기 사용량은 탄소 배출량과 상관관계가 있으며 시간 및 위치에 구애받지 않습니다.

또한 GPU는 종종 각 코어가 각 시점에서 소비하는 전력량을 보고하므로 AI 결과를 생성하여 소비되는 총 전력량을 쉽게 추정할 수 있습니다. 그럼에도 불구하고 이 측정은 하드웨어에 따라 다르며 결과적으로 서로 다른 모델 간의 공정한 비교를 허용하지 않습니다.

경과된 실시간 AI 결과를 생성하기 위한 총 실행 시간은 다른 모든 조건이 동일할 때 더 빠른 모델이 계산 작업을 덜 수행하므로 효율성을 위한 자연스러운 척도입니다. 그럼에도 불구하고 이 측정은 기본 하드웨어, 동일한 시스템에서 실행되는 다른 작업 및 사용된 코어 수와 같은 요인의 영향을 많이 받습니다.

이러한 요소는 서로 다른 모델 간의 비교와 하드웨어 개선에서 모델링 기여의 분리를 방해합니다.

매개변수 의 수 효율의 또 다른 일반적인 척도는 모델에서 사용되는 매개변수(학습 가능 또는 전체)의 수입니다. 런타임과 마찬가지로 이 측정값은 작업량과 상관관계가 있습니다. 위에서 설명한 다른 조치와 달리 기본 하드웨어에 의존하지 않습니다. 또한 이 측정값은 모델이 소비하는 메모리 양과도 높은 상관관계가 있습니다. 그럼에도 불구하고 서로 다른 알고리즘은 매개변수를 다르게 사용합니다. 예를 들어 모델을 더 깊게 만들거나 더 넓게 만듭니다. 결과적으로 유사한 수의 매개 변수를 가진 다른 모델은 종종 다른 양의 작업을 수행합니다.

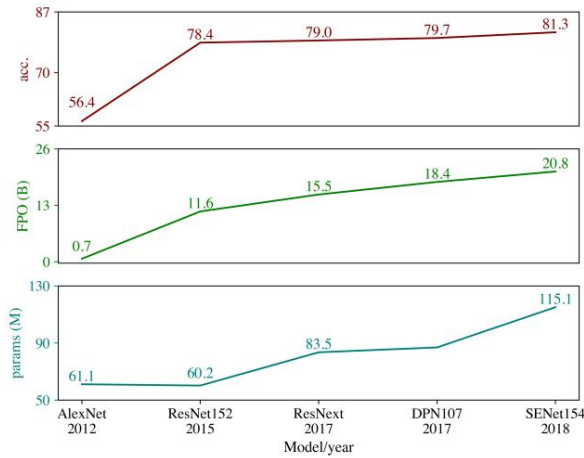
FPO 결과를 생성하는 데 필요한 FPO(부동 소수점 연산)의 총 수를 보고하는 것이 구체적인 방법입니다.¹³ FPO는 계산 프로세스에서 수행되는 작업량에 대한 추정치를 제공합니다. ADD 와 MUL 이라는 두 가지 기본 작업에 대한 비용을 정의하여 분석적으로 계산됩니다. 이러한 작업을 기반으로 기계 학습 추상 작업(예: tanh 작업, 행렬 곱셈, 컨볼루션 작업 또는 BERT 모델)의 FPO 비용을 이 두 작업의 재귀 함수로 계산할 수 있습니다. FPO는 과거에 모델의 에너지 발자국을 정량화하는 데 사용되었지만^[27, 43, 12, 42], AI에서는 널리 채택되지 않았습니다.

FPO에는 몇 가지 매력적인 속성이 있습니다. 첫째, 모델의 특정 인스턴스를 실행할 때 실행 중인 시스템이 수행한 작업량을 직접 계산하므로 소비되는 에너지의 양과 연결됩니다. 둘째, FPO는 모델이 실행되는 하드웨어에 구애받지 않습니다. 이는 위에서 설명한 측정과 달리 서로 다른 접근 방식 간의 공정한 비교를 용이하게 합니다. 셋째, FPO는 모델의 실행 시간과 밀접한 관련이 있습니다^[4]. 점근적 런타임과 달리 FPO는 각 시간 단계에서 수행되는 작업량도 고려합니다.

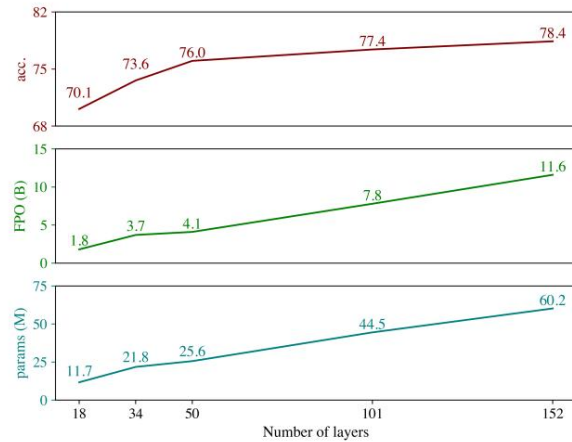
다양한 신경망 라이브러리¹⁴에서 FPO를 계산하기 위한 여러 패키지가 존재 하지만, 그 중 어느 것도 모든 최신 AI 모델을 구성하는 데 필요한 모든 빌딩 블록을 포함하지 않습니다. 우리는 신경망 라이브러리 빌더가 이러한 기능을 직접 구현하도록 권장합니다.

¹³ 부동 소수점 연산은 종종 FLOP(s)라고 하지만 이 용어는 고유하게 정의되지 않습니다^[12]. 혼동을 피하기 위해 용어를 사용합니다. FPO.

¹⁴E.g., <https://github.com/Swall0w/torchstat> ; <https://github.com/Lyken17/pytorch-OpCounter>



(a) 다른 모델.



(b) ResNet 모델의 다른 계층.

그림 4: FPO 증가로 인해 물체 감지 상위 1위 정확도에 대한 반환이 감소합니다. 플롯(아래에서 위로): 모델 매개변수(백만 단위), FPO(십억 단위), ImageNet의 상위 1위 정확도. (4a): 다른 모델: AlexNet[20], ResNet[14], ResNext[47], DPN107[5], SENet154[17]. (4b): ResNet 모델의 다양한 크기(레이어 수로 측정) 비교[14].

토론 효율적인 기계 학습 접근 방식은 연구 커뮤니티에서 주목을 받았지만 일반적으로 친환경적이지 않습니다. 예를 들어, 컴퓨터 비전 커뮤니티에서 상당한 양의 작업이 효율적인 추론을 다루었습니다. 이는 자율 주행 자동차[24, 31, 22]와 같은 애플리케이션을 위한 이미지의 실시간 처리 또는 다음과 같은 장치에 모델을 배치하는 데 필요합니다. 휴대폰으로 [16, 34]. 이러한 접근 방식의 대부분은 효율적인 모델 추론[32, 50, 12], 15를 목표로 하므로 단일 예제 처리 비용만 최소화하고 섹션 2.16 에서 논의된 다른 두 가지 **빨간색** 관행은 무시합니다.

위의 예는 AI를 **녹색** 으로 만드는 경로가 사용 방법에 달려 있음을 나타냅니다. 새 모델을 개발할 때 대부분의 연구 프로세스에는 훈련 세트에서 많은 모델 변형을 훈련하고 작은 개발 세트에서 추론을 수행하는 것이 포함됩니다. 이러한 환경에서 보다 효율적인 교육 절차는 더 큰 비용 절감으로 이어질 수 있는 반면 생산 환경에서는 보다 효율적인 추론이 더 중요할 수 있습니다. 우리는 다른 분야를 발전시키기 위해 일부 영역을 희생하지 않는 전산 절약에 대한 전체론적 관점을 지지합니다.

FPO에는 몇 가지 제한 사항이 있습니다. 첫째, 모델의 전력 소비를 목표로 하고 모델의 메모리 소비와 같이 종종 추가 에너지 및 금전적 비용으로 이어질 수 있는 연구원의 다른 잠재적인 제한 요소를 무시합니다[24]. 둘째, 모델이 수행하는 작업의 양은 모델 구현에 따라 크게 달라집니다. 동일한 모델의 두 가지 다른 구현이 매우 다른 처리 작업량을 초래할 수 있기 때문입니다. 모델링 기여에 중점을 두기 때문에 AI 커뮤니티는 전통적으로 모델 구현의 품질이나 효율성을 무시해 왔습니다. AI 커뮤니티에서 인정합니다.

3.2 기존 모델의 FPO 비용 작업량 보고의 중요성을 설

명하기 위해 여러 기존 모델에 대한 FPO 비용을 제시합니다. 18 그림 4a는 여러 주요 객체 인식 모델의 매개변수 수와 FPO를 보여줍니다.

15최근 일부 작업은 효율적인 교육을 목표로 했습니다[7].

16 사실 더 작은 모델을 만들면 실행 시간이 길어지는 경우가 많기 때문에 서로 다른 추세를 완화하는 것은 어려울 수 있습니다[44].

17 최종 예측에 대한 이러한 배타적인 초점은 **Red AI의 또 다른 증상이라고 생각합니다.**

18 이 숫자는 추론당 FPO, 즉 단일 예제를 처리하는 데 필요한 작업을 나타냅니다.

ImageNet 데이터세트의 성능 [6]. 19 몇 가지 추세를 관찰할 수 있습니다. 첫째, 섹션 2에서 논의한 바와 같이 모델은 시간이 지남에 따라 더 비싸지지만 FPO의 증가는 유사한 성능 향상으로 이어지지 않습니다. 예를 들어, ResNet과 ResNext(그래프의 두 번째 및 세 번째 지점) 사이의 FPO가 거의 35% 증가하면 상위 1위 정확도가 0.5% 향상되었습니다. 모델 작업의 다른 증가 효과를 고려할 때 유사한 패턴이 관찰됩니다. 둘째, 모델 매개변수의 수가 전체 내용을 말해주지는 않습니다. AlexNet(그래프의 첫 번째 지점)에는 실제로 ResNet(두 번째 지점)보다 더 많은 매개변수가 있지만 FPO는 훨씬 적고 정확도도 훨씬 낮습니다.

그림 4b는 단일 객체 인식 모델인 ResNet[14]에 대한 동일한 분석을 보여주며, 다른 버전의 모델을 다른 수의 레이어와 비교합니다. 이는 크기(따라서 FPO 비용)를 제외하고 아키텍처가 동일하기 때문에 서로 다른 모델 간에 통제된 비교를 생성합니다. 다시 한 번 동일한 추세를 확인할 수 있습니다. FPO 비용이 크게 증가해도 성능이 크게 향상되지는 않습니다.

3.3 그린 AI를 촉진하기 위한 추가적인 방법

최종 보고된 수의 FPO 비용을 보고하는 것 외에도 연구원은 교육 중에 관찰된 새싹 획득/정확도 곡선을 보고할 것을 권장합니다. 최근 논문[9]에서 우리는 주어진 작업에서 더 나은 성능의 모델을 선택하는 것이 모델 개발 중에 사용 가능한 컴퓨팅 양에 크게 의존한다는 것을 관찰했습니다. 우리는 주어진 예산의 함수로서 모델의 예상 최고 검증 성능을 계산하는 방법을 도입했습니다. 우리는 이 곡선을 보고하면 사용자가 모델 선택에 대해 더 현명한 결정을 내리고 다양한 접근 방식의 안정성을 강조할 수 있다고 주장합니다.

우리는 검토자에게 최신 기술을 엄격하게 개선하지는 않지만 효율성과 같은 다른 이점이 있는 기여를 인식하고 평가하도록 조언함으로써 주요 AI 컨퍼런스에서 효율성을 공식적인 기여로 만들 것을 옹호합니다. 마지막으로, 우리는 사전 훈련된 모델을 공개적으로 공개하는 추세가 녹색 성공이라는 점에 주목하고 조직이 다른 사람들이 모델을 재훈련하는 비용을 절약하기 위해 모델을 계속 출시하도록 권장하고 싶습니다.

4 관련 업무

최근 연구에서는 심층 NLP 모델 훈련의 탄소 배출량을 분석했으며 [40] 계산 비용이 많이 드는 실험이 환경 및 경제에 큰 영향을 미칠 수 있다고 결론지었습니다. 이러한 대규모 예산을 사용하는 현대 실험에서는 많은 연구자(특히 학계의 연구자)가 많은 유명 분야에서 작업할 자원이 부족합니다. 계산적으로 효율적인 접근법에 부여된 가치 증가는 보다 다양한 그룹의 연구 기여를 허용할 것입니다. 우리는 [40]의 결론이 장기적인 추세의 결과이며 NLP 내에서 격리되지 않고 기계 학습 전반에 걸쳐 사실임을 강조합니다.

일부 회사는 탄소 크레딧을 구매하여 전기 사용량을 상쇄하지만 크레딧 구매가 에너지를 덜 사용하는 것만큼 효과적인지는 확실하지 않습니다. 또한 탄소 배출권 구매는 자발적입니다. Google 클라우드20 및 Microsoft Azure21은 사용된 에너지를 상쇄하기 위해 탄소 배출권을 구매하지만 Amazon의 AWS22 (가장 큰 클라우드 컴퓨팅 플랫폼23)는 전력 사용량의 50%만 재생 가능 에너지로 처리했습니다.

최첨단 성능을 개선하려는 노력은 모델 개발 및 하이퍼파라미터 튜닝을 위한 많은 실험을 실행한 후 단일 최상의 결과를 보고하는 데 연구 커뮤니티의 관심을 집중시켰습니다. 이러한 실험을 완전히 보고하지 않으면 미래의 연구자가 결과를 재현하거나 확장하는 데 얼마나 많은 노력이 필요한지 이해하지 못합니다[9].

우리의 초점은 기계 학습 커뮤니티의 효율성을 개선하는 데 있지만 기계 학습은 기후 변화와 같은 영역에서 작업을 위한 도구로도 사용할 수 있습니다. 예를 들어 기계 학습은 시멘트 공장의 배출 감소[1] 및 동물 보호 결과 추적[11]에 사용되었으며 산물 관리에 유용할 것으로 예상됩니다[33]. 의심할 여지없이 이것들은 기계 학습의 중요한 응용 프로그램입니다. 우리는 그것들이 이 백서의 내용과 직교한다는 것을 알고 있습니다.

19<https://github.com/sovrasov/flops-counter.pytorch>에서 가져온 숫자 20<https://cloud.google.com/sustainability/> 21<https://www.microsoft.com/en-us/environment/carbon> 22<https://aws.amazon.com/about-aws/sustainability/> 23<https://tinyurl.com/y2kob969>

5. 결론

Green AI의 비전은 Red AI의 포용성 문제를 극복하는 데 도움이 되는 많은 흥미로운 연구 방향을 제기합니다. 진행률은 최소한의 성능 저하로 계산 비용을 줄이거나 더 효율적인 방법이 발견되면 성능을 향상시킬 것입니다. 또한 Green AI는 두뇌가 매우 효율적이기 때문에 인자적으로 더 그럴듯한 방향으로 우리를 움직일 수 있을 것 같습니다.

우리는 Green AI를 배타적인 의무가 아닌 귀중한 옵션으로 보고 있음을 다시 한 번 강조하는 것이 중요합니다. 물론 Green AI와 Red AI 모두 기여할 기여가 있습니다. 우리는 Green AI의 이점을 강조하고 효율성의 표준 척도를 옹호함으로써 Green AI의 보급률을 높이고자 합니다. 아래에서 우리는 몇 가지 중요한 녹색 연구 방향을 가리키고 몇 가지 공개 질문을 강조합니다.

건물 공간 또는 시간 효율적인 모델에 대한 연구는 종종 소형 장치(예: 전화)에 모델을 맞추거나 시각 장애인을 위한 이미지 캡션과 같이 실시간으로 예제를 처리할 수 있을 만큼 빠릅니다(섹션 3.1 참조).

일부 최신 모델은 단일 GPU에도 맞지 않습니다(섹션 2 참조). 여기서 우리는 훨씬 더 광범위한 접근 방식을 주장합니다.

데이터 효율성은 수년에 걸쳐 상당한 관심을 받아 왔습니다[35, 19]. 비전 및 NLP에 대한 최신 연구에는 먼저 대규모 "원시"(주석이 없는) 데이터에 대한 모델을 사전 교육한 다음 감독 학습을 통해 관심 있는 작업으로 미세 조정하는 작업이 포함됩니다. 이 영역에서 강력한 결과는 더 적은 수의 교육 예제 또는 더 적은 기울기 단계로 기준선과 유사한 성능을 달성하는 경우가 많습니다. 가장 최근 작업은 미세 조정 데이터[29]를 다루었지만 사전 훈련 효율성도 중요합니다. 두 경우 모두 이 영역에서 개선할 수 있는 간단한 기술 중 하나는 다른 양의 교육 데이터로 성능을 보고하는 것입니다. 예를 들어, 1,000만, 1억, 10억, 100억 개의 토큰에 대해 훈련된 컨텍스트 임베딩 모델의 보고 성능은 가장 작은 데이터 크기에서 먼저 비교할 수 있으므로 새 모델을 더 빠르게 개발할 수 있습니다. 여기서 연구는 교육 비용을 줄이는 것뿐만 아니라 리소스가 적은 언어 또는 역사적 도메인과 같은 영역에서 더 많은 데이터를 생성하는 것이 매우 어렵기 때문에 가치가 있으므로 진행하려면 사용 가능한 것을 보다 효율적으로 사용해야 합니다.

마지막으로, 최종 결과를 얻기 위해 실행되는 총 실험 수는 종종 과소 보고되고 과소 논의됩니다[9]. 보고된 실험 결과에 포함된 하이퍼파라미터 검색, 아키텍처 평가 및 제거에 대한 전체 보고에 대해 연구원이 가지고 있는 몇 가지 사례는 커뮤니티를 놀라게 했습니다[40]. 주어진 성능 수준에 도달하는 데 필요한 계산 비용을 줄일 수 있는 많은 하이퍼파라미터 최적화 알고리즘이 존재하지만 [3, 10] 여기서 간단한 개선은 큰 영향을 미칠 수 있습니다. 예를 들어 성능이 확실히 떨어지는 모델에 대한 교육을 초기에 중단하면 상당한 절감 효과를 얻을 수 있습니다[21].

참조

[1] Prabal Acharyya, Sean D Rosario, Roey Flor, Ritvik Joshi, Dian Li, Roberto Linares 및 Hongbao Zhang.

연료 소비 및 배출 감소를 위한 시멘트 공장의 자동 조종 장치, 2019. 기후 변화에 관한 ICML 워크숍.

[2] 다리오 아모데이와 대니 에르난데즈. AI 및 컴퓨팅, 2018. 블로그 게시물.

[3] James S. Bergstra, Remi Bardenet, Yoshua Bengio 및 Bal azs K. 예. 하이퍼 매개변수 최적화를 위한 알고리즘입니다. 프로세스에서 NeurIPS, 2011.

[4] Alfredo Canziani, Adam Paszke, Eugenio Culurciello. 실제 응용을 위한 심층 신경망 모델 분석. 프로세스에서 ISCAS, 2017.

[5] Yunpeng Chen, Jianan Li, Huaxin Xiao, Xiaojie Jin, Shuicheng Yan 및 Jiashi Feng. 이중 경로 네트워크. 프로세스에서 NeurIPS, 2017.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li 및 Li Fei-Fei. ImageNet: 대규모 계층 구조 이미지 데이터베이스. 프로세스에서 CVPR, 2009.

[7] 팀 데트머스와 루크 제틀모이어. 처음부터 희소 네트워크: 성능 저하 없이 더 빠른 훈련, 2019. arXiv:1907.04840.

- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: 심층 양방향 사전 훈련 언어 이해를 위한 변환기. 프로세스에서 NAACL, 2019.
- [9] Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, Noah A. Smith. 작업 표시: 개선된 실험 결과 보고. 프로세스에서 EMNLP, 2019.
- [10] 제시 닛지, 케빈 제이미슨, 노아 A. 스미스. 개방 루프 하이퍼파라미터 최적화 및 결정 포인트 프로세스. 프로세스에서 AutoML, 2017.
- [11] Clement Duhart, Gershon Dublon, Brian Mayton, Glorianna Davenport, Joseph A. Paradiso. 야생 동물 보호 및 복원 노력을 위한 딥 러닝, 2019. 기후 변화에 관한 ICML 워크숍.
- [12] Ariel Gordon, Elad Eban, Ofir Nachum, Bo Chen, Hao Wu, Tien-Ju Yang, Edward Choi. MorphNet: 딥 네트워크의 빠르고 간단한 자원 제약 구조 학습. 프로세스에서 CVPR, 2018.
- [13] Alon Halevy, Peter Norvig, Fernando Pereira. 데이터의 비합리적인 효율성. IEEE 지능형 시스템, 24:8-12, 2009.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. 이미지 인식을 위한 심층 잔차 학습. ~ 안에 절차 CVPR, 2016.
- [15] Sepp Hochreiter와 Jürgen Schmidhuber. 장단기 기억. 신경 계산, 9(8):1735-1780, 1997.
- [16] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto 및 Hartwig Adam. MobileNets: 모바일 비전 애플리케이션을 위한 효율적인 컨벌루션 신경망, 2017. arXiv:1704.04861.
- [17] Jie Hu, Li Shen, Gang Sun. 압축 및 여기 네트워크. 프로세스에서 CVPR, 2018.
- [18] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama 및 Kevin Murphy. 최신 컨벌루션 객체 감지기의 속도/정확도 트레이드오프. 프로세스에서 CVPR, 2017.
- [19] Sanket Kamthekar와 Marc Peter Deisenroth. 확률적 모델 사전을 통한 데이터 효율적인 강화 학습 지시적 통제. 프로세스에서 AISTATS, 2018.
- [20] Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton. 깊은 컨벌루션 신경망을 사용한 이미지넷 분류 네트워크. 프로세스에서 NeurIPS, 2012.
- [21] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, Ameet Talwalkar. 하이퍼밴드: 산적 하이퍼파라미터 최적화를 위한 기반 구성 평가. 프로세스에서 ICLR, 2017.
- [22] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. San. Ssd: 단일 샷 멀티박스 감지기. 프로세스에서 ECCV, 2016.
- [23] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov. RoBERTa: 강력하게 최적화된 bert 사전 훈련 접근 방식, 2019. arXiv:1907.11692.
- [24] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, Jian Sun. ShuffleNet V2: 효율적인 cnn 아키텍처 설계를 위한 실용적인 지침. 프로세스에서 ECCV, 2018.
- [25] Dhruv Mahajan, Ross Girshick, Vignesh Ramanathan, Kaiming He, Manohar Paluri, Yixuan Li, Ashwin Bharambe 및 Laurens van der Maaten. 약하게 감독된 사전 훈련의 한계를 탐색합니다. 프로세스에서 ECCV, 2018.

- [26] 가버 엘리스, 크리스 다이어, 필 블런션. 신경 언어 모델의 최신 평가 기술에 대해. 프로세스에서 EMNLP, 2018.
- [27] Pavlo Molchanov, Stephen Tyree, Tero Karras, Timo Aila 및 Jan Kautz. 컨벌루션 신경망 가지치기 리소스 효율적인 추론을 위해. 프로세스에서 ICLR, 2017.
- [28] 고든 E. 무어. 1965년, 집적 회로에 더 많은 부품을 집어넣음.
- [29] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettl 모이어 심층적으로 문맥화된 단어 표현. 프로세스에서 NAACL, 2018.
- [30] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever. 언어 모델은 감독되지 않은 멀티태스킹 학습자입니다, 2019. OpenAI 블로그.
- [31] Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, Ali Farhadi. Xnor-net: 이진 길쌈 신경망을 사용한 Imagenet 분류. 프로세스에서 ECCV, 2016.
- [32] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: 통합된 실시간 객체 발각. 프로세스에서 CVPR, 2016.
- [33] David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, An Drew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Muktavilli, Konrad P. Kording, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, Yoshua Bengio. 기계 학습을 통한 기후 변화 문제 해결, 2019. arXiv:1905.12616.
- [34] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov 및 Liang-Chieh Chen. 모바일넷V2: 반전 잔차 및 선형 병목 현상. 프로세스에서 CVPR, 2018.
- [35] 로이 슈워츠, 샘 톰슨, 노아 A. 스미스. SoPa: CNN, RNN 및 가중 유한 상태 머신을 연결합니다. 프로세스에서 ACL, 2018.
- [36] Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Nieves, Terah Lyons, John Etchemendy, Z Bauer. AI 인덱스 2018 연례 보고서. AI 인덱스 운영위원회, 인간 중심 AI 이니셔티브, 스탠포드 대학. <http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf>, 202018, 2018 에서 확인 가능 .
- [37] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever , Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel 및 Demis Hassabis. 심층 신경망 및 트리 검색으로 바둑 게임 마스터하기 네이처, 529(7587):484, 2016.
- [38] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dhharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan 및 Demis Hassabis. 일반 강화 학습 알고리즘을 사용한 셀프 플레이로 체스와 장기 마스터하기, 2017. arXiv:1712.01815.
- [39] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, 토르 그레펠, 데미스 하사비스. 인간의 지식 없이 바둑 게임을 마스터합니다. 자연, 550(7676):354, 2017.
- [40] Emma Strubell, Ananya Ganesh, Andrew McCallum. 딥 러닝을 위한 에너지 및 정책 고려 사항 자연어 처리 프로세스에서 ACL, 2019.

- [41] Chen Sun, Abhinav Shrivastava, Saurabh Singh, Abhinav Gupta. 딥러닝 시대에 데이터의 비합리적인 효율성을 재검토하다. 프로세스에서 ICCV, 2017.
- [42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, 그리고 일리아 폴로수힌. 주의가 필요합니다. 프로세스에서 NeurIPS, 2017.
- [43] Tom Veniat 및 Ludovic Denoyer. 예산이 책정된 슈퍼 네트워크 작업으로 시간/메모리 효율적인 딥 아키텍처를 학습합니다. 프로세스에서 CVPR, 2018.
- [44] Aaron Walsman, Yonatan Bisk, Saadia Gabriel, Dipendra Misra, Yoav Artzi, 최예진, Dieter Fox. 일찍 목표 지향 로봇 비전을 위한 융합. 프로세스에서 IROS, 2019.
- [45] Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, Samuel R. Bowman. SuperGLUE: 범용 언어 이해 시스템에 대한 보다 엄격한 벤치마크, 2019. arXiv:1905.00537.
- [46] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, Samuel R. Bowman. GLUE: 자연어 이해를 위한 멀티태스킹 벤치마크 및 분석 플랫폼입니다. 프로세스에서 ICLR, 2019.
- [47] Saining Xie, Ross Girshick, Piotr Dollar, Zhuowen Tu 및 Kaiming He. 심층 신경망에 대한 집계된 잔차 변환. 프로세스에서 CVPR, 2017.
- [48] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov 및 Quoc V. Le. XLNet: 언어 이해를 위한 일반화된 자기회귀 사전 훈련, 2019. arXiv:1906.08237.
- [49] Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, 최예진. 신경 가짜 뉴스에 대한 방어, 2019. arXiv:1905.12616.
- [50] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, Jian Sun. ShuffleNet: 매우 효율적인 컨볼루션 모바일 장치용 신경망. 프로세스에서 CVPR, 2018.
- [51] Barret Zoph 및 Quoc V. Le. 강화 학습을 통한 신경 구조 검색. 프로세스에서 ICLR, 2017.