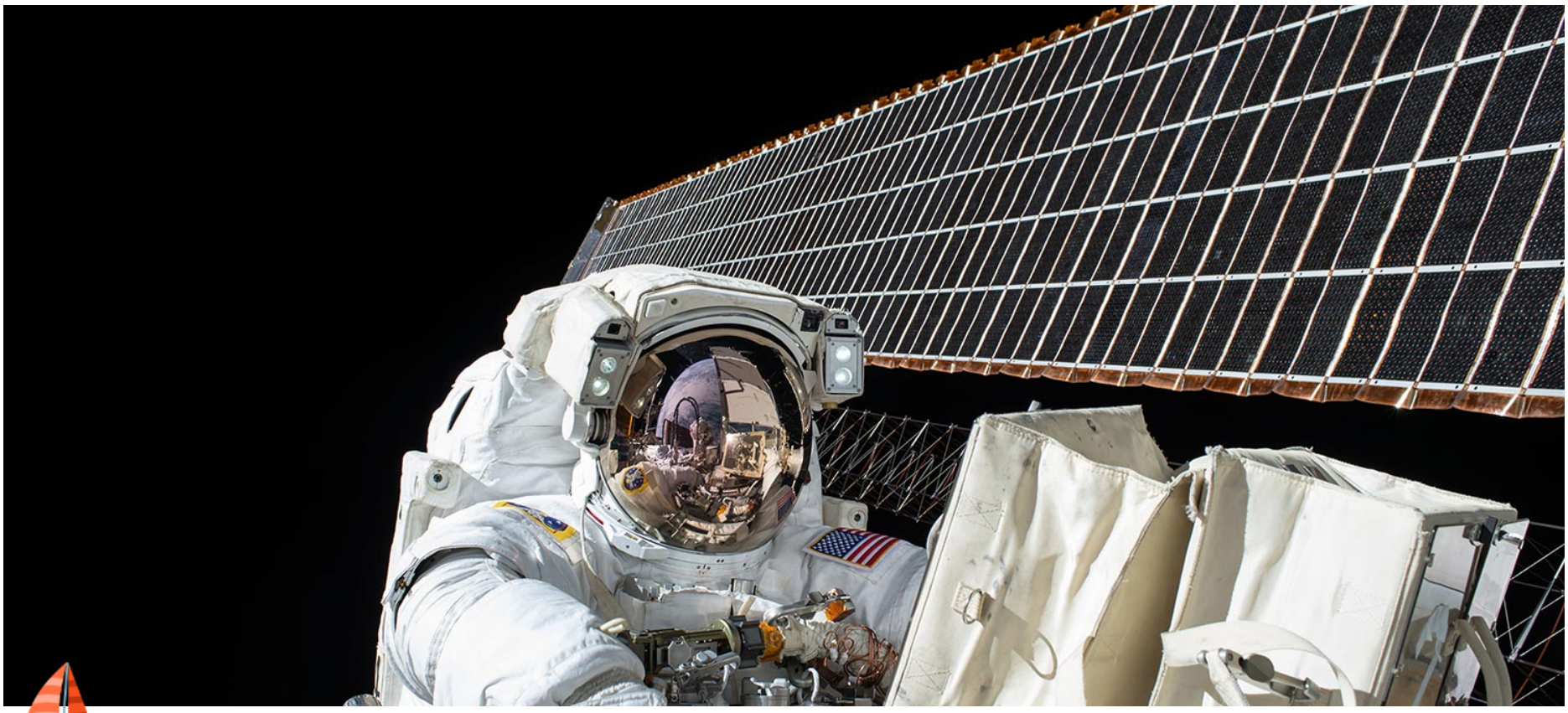


Taskonomy: Disentangling Task Transfer Learning

ABSTRACT

컴퓨터 비젼 테스크는 서로 관련이 있을까? 예를 들어 surface normal task는 depth image와 관련이 있을까요? 직관적으로 관련이 있을 것으로 보입니다. 그리고 이는 비젼 테스크간 구조에서 관련이 있을 것 같습니다. 이러한 관련은 전이 학습에서 상당한 가치가 있습니다. 관련이 있다면 학습한 것을 재사용하며 복잡성을 줄일 수 있습니다.

우리는 비젼 테스크에서의 관계 구조를 보여주고자 합니다. 이는 전이 학습에서 이루어지며 2D, 2.5D, 3D그리고 semantic segmentation과 같은 26가지의 테스크를 살펴봅니다. 테스크간 관계를 파악한다면 학습에 필요한 라벨값을 확기적으로 줄일 수 있을 것입니다. 우리는 10개의 테스크에 대해서 전이 학습이 이전 단독 학습보다 3분의 2정도의 데이터가 필요한 것을 증명했습니다. 우리는 테스크 관 관계를 볼 수 있는 방식과 유저들이 학습하는데 고려하는데 볼법한 taxonomical 구조를 제시합니다.

Introduction

객체 인식, 깊이 추정, 가장자리 탐지, 포즈 추정 등 여러 비젼 테스크는 유용합니다. 이들 중 몇몇은 확실한 관계를 가지고 있습니다. 우리는 surface normal와 깊이 추정은 관련이 있고 vanising point는 방향 탐지에 유용하다고 생각합니다. 다른 비젼 테스크에서도 비슷한 관련이 있다고 여겨집니다.

컴퓨터 비젼 분야에서는 이러한 관계에 대해서 명시적으로 증명하지는 않았습니다. 우리는 기계 학습 분야에서 많은 진전을 이루었고 트레이닝 데이터를 학습하여 복잡한 매핑을 찾아내어 예측을 진행합니다. 이는 supervised-learning 지도 학습이라 하며 한 테스크에 대해서만 진행되기도 합니다. **이런한 단독 학습은 새로운 테스크에 대해 처음부터 다시 학습 시켜야 한다는 단점이 존재하며 테스크 마다의 많은 데이터를 요구하게 됩니다.**

테스크에서의 관계를 이용한다면 모델은 적은 라벨과 적은 연산량을 요구하며 좋은 방향으로 학습하게 됩니다. 통합 모델은 다양한 테스크를 해결할 수 있는 효율적이고 전반적인 모델입니다. 그러나 테스크에 대한 이해는 아직 잘 모릅니다. 테스크 관계증명은 중요하며 학습 및 최적화를 통하여 이를 찾아내는 것은 복잡합니다.

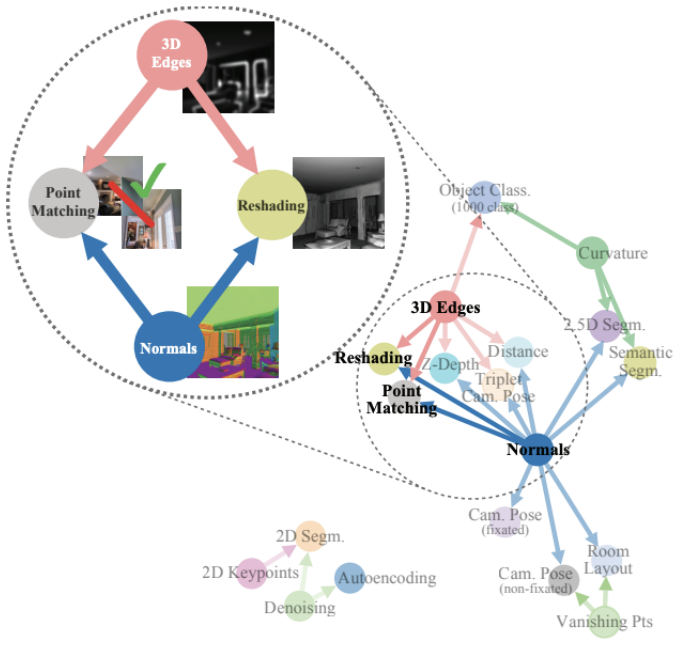


Figure 1: **A sample task structure discovered by the computational task taxonomy (*taskonomy*).** It found that, for instance, by combining the learned features of a surface normal estimator and occlusion edge detector, good networks for reshading and point matching can be rapidly trained with little labeled data.

이 논문에서 우리는 테스크간 관계를 찾아내는 모델을 제시하며 비젼 테스크간 관계 매핑을 제공합니다. 이러한 모델은 인공 신경망을 통해서 밝혀냅니다. 순방향 네트워크에서 각각의 레이어는 인풋 데이터를 통하여 추상적인 표현을 제공하며 이는 다시 아웃풋으로 도출됩니다. 그러나 이러한 representation은 다른 테스크를 해결하는데도 유용할 수 있습니다.

우리는 하나의 테스크를 학습한데 쓰인 representation이 다른 테스크의 결과물을 도출하는데 얼마나 유용한지에 대해 행렬을 계산합니다. 이러한 방식은 단독 학습보다 다른 테스크에서 학습한 것을 기반으로 하기 때문에 해당 테스크에서 데이터를 덜 요구할 수 있게 됩니다.

representation-based와 fully-computational기반이므로 가정을 피할 수 있습니다. 가정이 없는 것은 사람의 의견이나 분석적인 지식이 들어가지 않기 때문에 중요한 점입니다. 예를 들어 우리는 깊이 추정이 표면 추출과 같은 곳에 도움이 될 것이라 여길 수 있을지라도 사전 가정이 없으면 컴퓨터적 계산에 의해 더욱 좋은 방향을 찾아냅니다.

RELATED WORK

요즘 사용하는 기술은 튜링이 주장했던 학습방식인 이전 스테이지에서 결과물을 토대로 학습해서 나아가는 방식이 발전되어 왔습니다. 우리는 이러한 구조를 찾고자 했습니다. 우리가 하고자 하는 것은 다양한 분야와 관련이 있습니다.

Self-supervised learning

자기 지도 학습은 상대적으로 싼 테스크를 통해 학습한 정보를 이용하여 관련이 있는 보다 비싼 테스크를 학습하는 방식입니다. 이는 source task를 수동적으로 넣어주어야 하는 단점이 존재합니다.

Unsupervised learning

비지도 학습은 테스크에 대해 라벨이 주어지지 않은 환경에서 인풋에서의 유사한 점을 찾아 간결한 representation을 표현하는데 목표를 둡니다.

Meta-learning

메타 러닝은 전통적인 학습 단계 보다 위에서 학습을 하는데 초점을 두고 있습니다. 학습 데이터셋 상관없이 보편적인 방법을 찾는데 사용됩니다. 예를 들어 강화학습, 최적화 등에서 사용됩니다.

Multi-task learning

멀티 테스크 러닝은 하나의 인풋을 가지고 여러 테스크에서 사용할 수 있도록 다양한 아웃풋을 학습하는 방식입니다.

Domain adaption

도메인 적용은 같은 테스크에서 특정 도메인에서 다른 도메인으로의 적용을 위해서 학습하는 방식입니다. 이도 전이 학습이라 할 수 있지만, 우리는 테스크에서 다른 테스크를 다루고 있습니다.

Learning Theoretic

이 접근은 위에서의 모든 방식을 약간씩 사용한 접근입니다. 일반화할 수 있는 성능을 보장하는데 초점을 두고 있습니다. 이는 모델과 테스크에 대해서 제한을 두어 다루기 힘든 계산을 피했습니다. 우리는 이러한 것과 유사하되 보다 실용적인 접근을 하고자 합니다.

METHOD

우리는 문제를 다음과 같이 정의합니다. 우리는 R 이라는 라벨 제한을 두고 (데이터 셋 양의 제한을 두고) $T = [t_1, ..., t_n]$ 테스크에 대해 전체적은 성능을 최대화하는데 목적을 둡니다. R이라는 데이터 셋 양은 사용 테스크에 따라 최대 양이 정해질 수 있습니다. V= S⋃T 이며 T는 우리가 풀고자 하는 테스크이고 S는 학습할 수 있는 Source입니다. 그러므로 T - T ⋂S는 우리가 학습할 수 없는 타켓만이 존재하며 S-T⋂S는 소스만이 존재합니다. 이는 직접적으로 해당 테스크를 해결할 수는 없지만, T의 성능을 올리기 위해서 사용가능합니다.

task taxonomy(테스크 분류, taskonomy)는 계산적으로 테스크 전이성을 포착하여 그래프로 표현합니다. 소스 테스크와 타겟 테스크의 사이 엣지는 관련성을 나타내며 이것의 가중치는 해당 작업의 예측입니다. 우리는 이러한 엣지를 전체적인 선택적 전이를 측정하기 위해 사용합니다. 분류는 지도학습 양, task 선택, 전이순서, 전이 함수의 표현에 따라 그래프와 파라미터를 표현합니다.

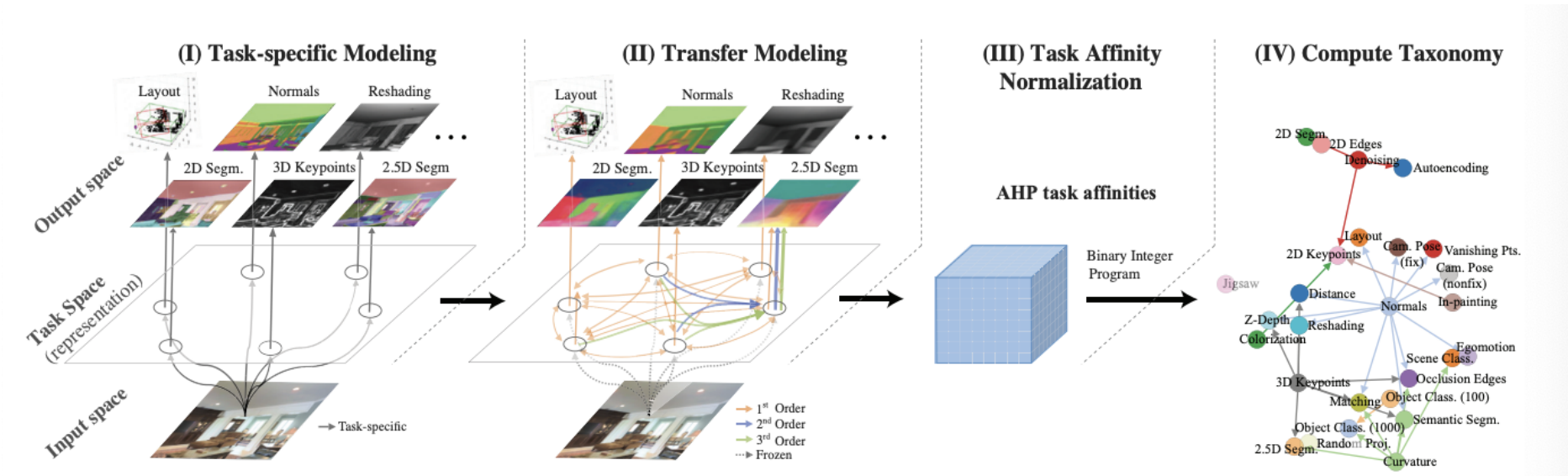


Figure 2: **Computational modeling of task relations and creating the taxonomy.** From left to right: I. Train task-specific networks. II. Train (first order and higher) transfer functions among tasks in a latent space. III. Get normalized transfer affinities using AHP (Analytic Hierarchy Process). IV. Find global transfer taxonomy using BIP (Binary Integer Program).

Taxonomy는 4가지 단계를 통해 만들어집니다.

1. 각각의 소스 테스크에서 network를 학습합니다.
2. 가능한 모든 소스와 타겟 테스크 간 전이 학습을 진행합니다. 이 때 인풋 소스로 여러가지를 순서에 따라 사용하여 진행합니다.
3. 전이 학습으로 부터 얻어진 테스크 선호 값은 정규화됩니다.
4. 전이 학습으로 최고로 최적화할 수 있는 그래프를 생성해냅니다.

Task Dictionary

우리의 테스크 맵핑은 26개의 테스크로 부터 생성되어집니다. 이는 테스크에 대한 표현 공간을 의미합니다. 우리는 테스크에서 다양한 표현 값을 얻게 됩니다. 우리는 이러한 표현 값이 전체 공간이 아닌 시각적인 밀집된 공간에서 샘플링된 값이라는 것을 알아야 합니다. 이러한 밀집된 공간에서 출발한 학습은 보다 유리할 수 있다는 가정입니다. 더 좋은 샘플링일 수록 더 좋은 일반화를 가집니다.

Dataset

우리는 이미지에서 모든 테스크에 대한 라벨 값이 있는 데이터 셋을 사용합니다. 같은 데이터 셋을 할 경우 여러 가지 고정된 환경에서 실험을 할 수 있기에 테스크 간 전이 학습에 대한 결과를 정확히 알 수 있습니다. 지금까지 이러한 데이터 세트는 없었기 때문에 우리는 600개의 건물에 대해서 4백만장의 이미지를 생성했습니다. 라벨 값을 주기 위해 여러 툴을 사용을 했고, 이는 사람이 라벨링 한 것에서 7%미만의 오류를 보였습니다.

Task-speciifc Modeling

우리는 각각의 테스크에 대해 지도 학습을 진행합니다. 테스크 특정 네트워크는 인코더와 디코더 구조를 가집니다. 인코더는 충분히 강력한 representation을 추출하고 디코더는 인코더보다는 작지만 좋은 성능을 내기 위해 충분합니다.

Transfer Modeling

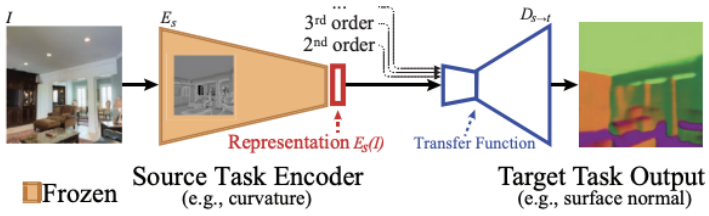
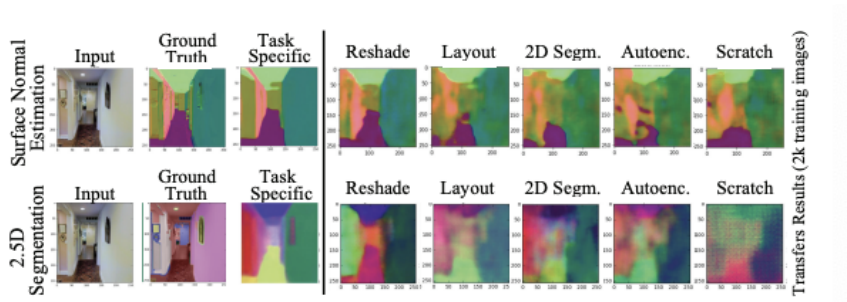


Figure 4: **Transfer Function**. A small readout function is trained to map representations of source task's frozen encoder to target task's labels. If order> 1, transfer function receives representations from multiple sources.

주어진 source로부터 학습한 representation을 통하여 target에 해당하는 task를 예측하기 위해 학습을 진행합니다. 이 때 target에 해당하는 task를 예측하기 어렵다면 source와 target task의 관련성은 적다고 볼 수 있습니다. 우리는 이처럼 모든 source와 target 조합을 이용하여 전이 학습을 진행합니다.



- Accessibility : 전이 학습을 성공시키기 위해 인코더는 이미지를 충분히 표현한 representation을 나타내야 합니다. 그리고 학습된 표현이 라벨값에 맞춰 잘 가는지 측정하기 위해 디코더는 작은 구조를 채택해야 하며 학습하는데 적은 데이터를 사용해야 하는 이유입니다.
- Higher-Order Transfoers : source tasks가 여러개를 사용하여 target task를 해결할 수 있습니다. 우리는 고차원의 전이 학습을 사용하게 됩니다. 고차원의 전이 학습 시 조합의 수는 정말 많습니다. Task가 25개 그리고 2개씩의 조합만을 선택한다 하여도 25C2 = 300개 이며 target에 대한 task 22개일 시 총 6600개의 조합을 시도해야 합니다. 물론 전체 조합을 고려하는 것이 좋은 결과를 내겠지만 이는 너무나 많은 연산을 소모하게 됩니다. 우리는 beam search를 사용합니다.

  beam search의 order를 k<5로 설정할 경우 우리는 저차원의 전이 학습을 사용했을 때 가장 성능이 좋은 5개를 선정하여 이 5개의 대한 순서 조합을 통한 실험을 진행합니다. 예를 들어 A라는 타겟 테스크가 있고, A에 대해서 S, X, K, L, G 테스크가 성능이 좋았을 때 이 5가지의 순서를 조합하여 진행합니다. 그리고 S → A 보다 K→S→A가 성능이 좋을 수도 있기 때문에 순서를 고려하며 조합을 하는 실험을 진행합니다.

Ordinal Normalization using Analytic Hierarchy Process (AHP)

각각의 테스크 간에 전이에 대한 관련성 행렬을 얻고자 합니다. 테스크에 대하여 Loss를 통하여 이를 산정하는 것은 테스크 간 공간에 따른 특이성이 존재하기 때문에 테스크 간 비교할 시 적합하지 않습니다. 그래서 적절한 정규화가 필요합니다. 간단한 방식은 스케일을 [0,1]의 단위로 변경하는 것입니다. 그러나 이 접근은 실제 출력물의 결과가 로스와는 다른 개선 속도가 있기 때문에 실패하였습니다. 로스와 퀄리티에 대한 관계는 알 수 없기 때문에 이러한 정규화는 효과가 없었습니다.

대신, 우리는 각각의 source i, j로 부터 학습을 진행한 결과값에 대하여 데이터 마다 어느 것이 성능이 더 좋았는지에 대해 비율을 산정하여 어떤 source가 더 좋은지 측정합니다. 최종적인 task간 affine 행렬은 아래와 같이 정의됩니다.

$$w'_{i,j} = \frac{\mathbb{E}_{I\in\mathcal{D}_{test}}[D_{s_i\to t}(I) > D_{s_j\to t}(I)]}{\mathbb{E}_{I\in\mathcal{D}_{test}}[D_{s_i\to t}(I) < D_{s_j\to t}(I)]}.$$

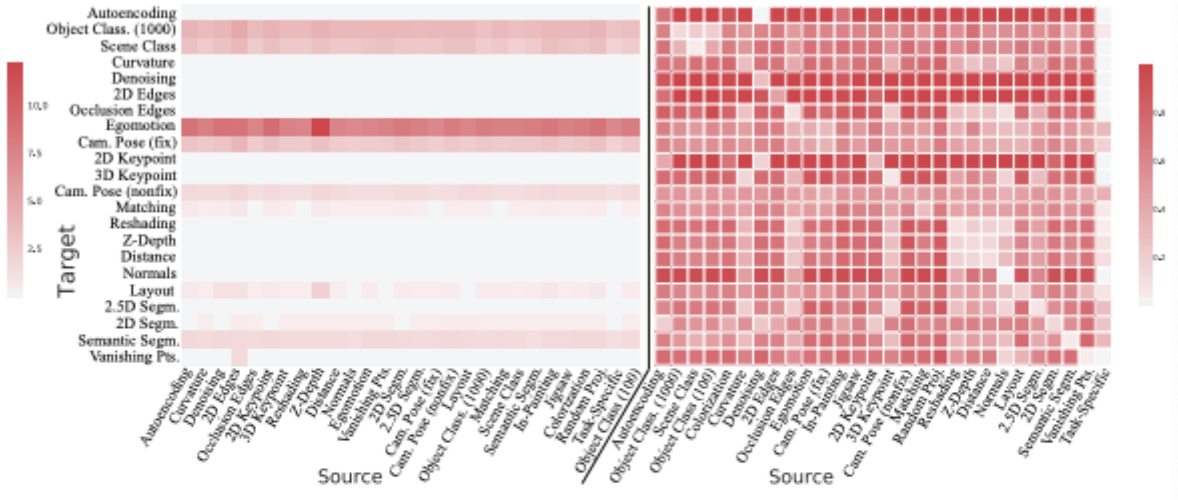


Figure 7: **First-order task affinity matrix** before (left) and after (right) Analytic Hierarchy Process (AHP) normalization. Lower means better transfered. For visualization, we use standard affinity-distance method $dist = e^{-\beta \cdot P}$ (where $\beta = 20$ and e is element-wise matrix exponential). See supplementary material for the full matrix with higher-order transfers.

각각의 소스 task를 타겟 task와 연산을 하고 모든 소스에 대해서 계산을 하고 이를 행을 쌓아 올리면 위와 같은 행렬 값이 나오게 됩니다. 이는 경영과학에서 흔히 사용되는 Analytic Hierarchy Process(AHP)에서 착안했다고 합니다. 위 사진의 오른쪽이 AHP를 적용한 행렬값입니다. 위에서의 그림은 저차원에 대해서만 수행을 한 것이고

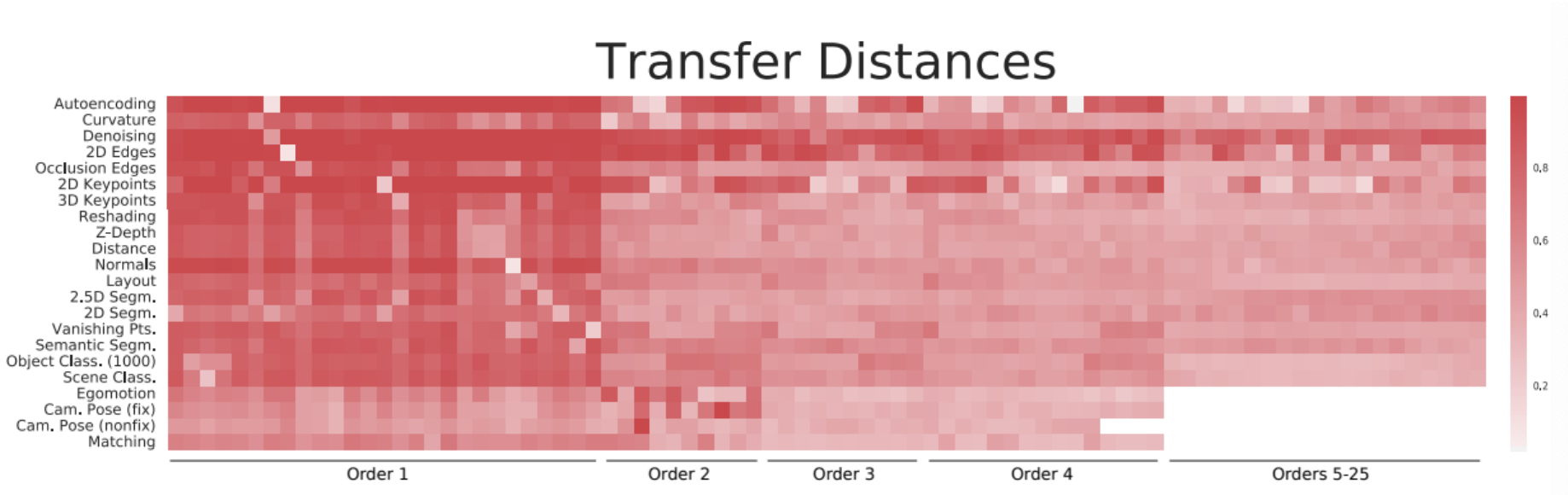


Figure 2: **Higher Order Distances.** Calculated task distances for all 16k transfers. The columns are sorted by transfer order and the orders are shown along the x-axis. Each row is a target task.

이는 여러개의 순서 쌍 조합을 통하여 나타낸 행렬 값입니다.

Computing the Global Taxonomy

위에서의 주어진 행렬을 이용하여 우리는 테스크에 대해서 전체적으로 성능을 극대화하며 지도 학습을 최소화하기 위해서 전이 학습 도안을 고안해야합니다. 이 문제는 노드와 엣지를 통해 그래프로 표현 가능합니다. 우리는 이것을 통해 특정 타겟 테스크의 성능을 높이기 위해 최적의 전이 학습 방식을 찾을 수 있습니다. 우리는 이것을 Boolean Integer Programming(BIP)을 사용합니다. 가장 성능이 좋은 순서를 이용하여 노드간의 연결을 키고 나머지는 끄는 방식을 통해 그래프를 제작합니다.

EXPERIMENTS

26개의 테스크 (4개는 source로만 사용), 26개의 특정 테스크 네트워크와 22*25 전이 학습 네트워크 그리고 22*25Ck의 고차원 전이 학습을 통하여 생성했습니다. 전이 학습 방식은 3000개 안쪽으로 되었으며 총 47,886 GPU 시간을 썼습니다. (약 2000일)

인코더 구조는 풀링 없이 ResNet-50을 기반입니다. 모든 전이 학습 구조는 얇은 2개의 convolution layers로 진행됩니다. 디코더의 구조는 task마다 상이한테 픽셀마다 예측이 필요한 경우는 15개의 conv layer를 사용합니다. 저차원의 task는 2~3개의 FC layer를 사용합니다. 각각의 지도학습된 네트워크가 얼마나 잘 representation을 만들었는지 확인하기 위하여 각각 네트워크의 성능은 아래와 같습니다.

Task	*avg*	*rand*	Task	*avg*	*rand*	Task	*avg*	*rand*
Denoising	100	99.9	Layout	99.6	89.1	Scene Class.	97.0	93.4
Autoenc.	100	99.8	2D Edges	100	99.9	Occ. Edges	100	95.4
Reshading	94.9	95.2	Pose (fix)	76.3	79.5	Pose (nonfix)	60.2	61.9
Inpainting	99.9	-	2D Segm.	97.7	95.7	2.5D Segm.	94.2	89.4
Curvature	78.7	93.4	Matching	86.8	84.6	Egomotion	67.5	72.3
Normals	99.4	99.5	Vanishing	99.5	96.4	2D Keypnt.	99.8	99.4
Z-Depth	92.3	91.1	Distance	92.4	92.1	3D Keypnt.	96.0	96.9
Mean	92.4	90.9						

Table 1: **Task-Specific Networks' Sanity:** Win rates vs. *random* (Gaussian) network representation readout and statistically informed guess *avg*.

Evaluation of Computed Taxonomies

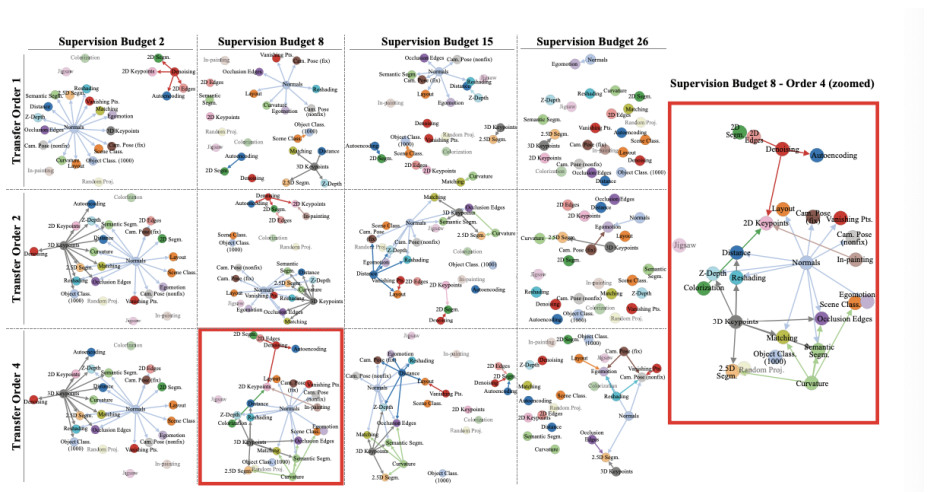


Figure 8: **Computed taxonomies** for solving 22 tasks given various supervision budgets (x-axes), and maximum allowed transfer orders (y-axes). One is magnified for better visibility. Nodes with incoming edges are target tasks, and the number of their incoming edges is the order of their chosen transfer function. Still transferring to some targets when tge budget is 26 (full budget) means certain transfers started performing better than their fully supervised task-specific counterpart. See the interactive solver website for color coding of the nodes by *Gain* and *Quality* metrics. Dimmed nodes are the source-only tasks, and thus, only participate in the taxonomy if found worthwhile by the BIP optimization to be one of the sources.

그래프의 나오는 시각화에 따라서 전이 학습 순서를 매칭하여 학습한다면 최고의 결과를 산출해낼 수 있게 된다.

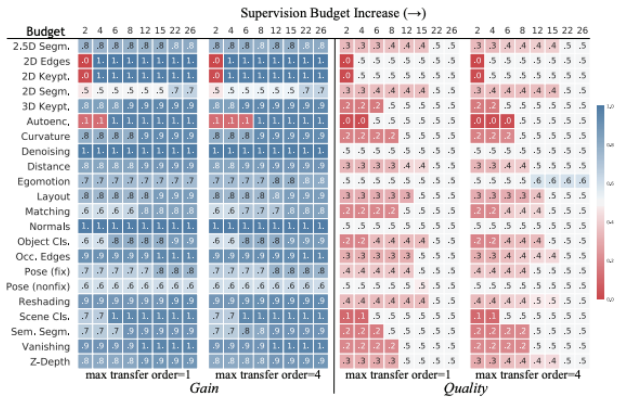


Figure 9: **Evaluation of taxonomy computed for solving the full task dictionary**. Gain (left) and Quality (right) values for each task using the policy suggested by the computed taxonomy, as the supervision budget increases(→). Shown for transfer orders 1 and 4.

GAIN : 특정 테스크 지도 학습 결과에 대해서 이긴 비율 (데이터 셋 1.6만)

QUALITY : 특정 테스크 지도 학습 결과에 대해서 이긴 비율 (데이터 셋 12만)

둘 다 0.5이상의 수치를 기록할 경우 전이 학습이 완전 지도 학습보다 성능이 좋은 것을 나타낸다. 또한 Source와 order의 수치가 증가할 수록 성능이 높아지는 경향이 있는데 이는 전이 학습을 많이 할수록 좋은 결과를 나타냄을 보인다. 그리고 Quality에서는 지도 학습의 성능이 더 좋은 경우가 많은데 이는 데이터 셋의 학습량이 많아서 이기 때문이다.

Significance Test of the Structure

Figure 11: **Structure Significance.** Our taxonomy compared with random transfer policies (random feasible taxonomies that use the maximum allowable supervision budget). Y-axis shows *Quality* or *Gain*, and X-axis is the supervision budget. Green and gray represent our taxonomy and random connectivities, respectively. Error bars denote 5^{th}–95^{th} percentiles.

Tasknomy에서 제시하는 전이 학습 방식을 따른 것이 초록색 선 랜덤으로 전이 학습을 진행한 것이 아래 있는 회색선이다. 이는, tasknomy에서 제시하는 순서가 더 좋음을 입증하며 테스크관 관계성이 있다는 것을 보여준다.

우리는 여러가지 방식을 고정해놓고 실험을 진행했다. 그러나 이러한 방식에서만 Tasknomy의 방법론이 통하는지 확인하기 위하여 아래와 같은 것들을 바꾸었으나 우리가 보여준 결과와 유사하게 나왔다.

I. architecture of task-specific networks

II. architecture of transfer function networks

III. amount of data available for training transfer networks

IV. datasets

V. data splits

VI. choice of dictionary

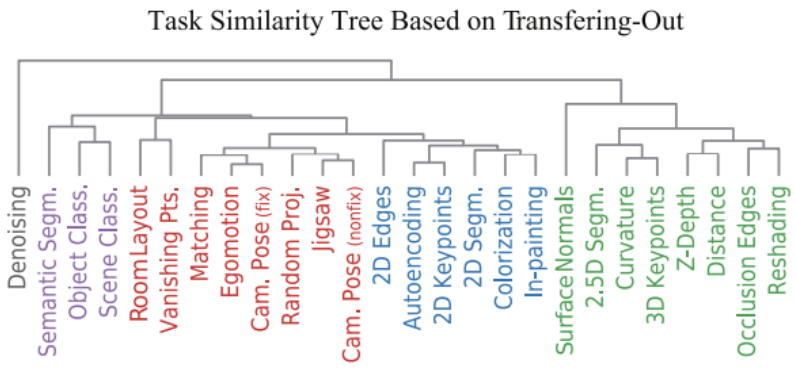


Figure 13: **Task Similarity Tree.** Agglomerative clustering of tasks based on their transferring-out patterns (i.e. using columns of normalized affinity matrix as task features). 3D, 2D, low dimensional geometric, and semantic tasks clustered together using a fully computational approach.

그래프가 아닌 나무기반 모형으로 표현을 하였고 이는 우리가 본 결과와 유사합니다.

Limitations and Discussion

Model Dependence

우리는 고정된 모델과 고정된 데이터 셋에 한정을 지어 실험을 수행하였기 때문에, 이는 다른 조건 상황에서도 유사한 결론이 나올지 확인해야 합니다.

Compositionality

우리가 수행한 테스크는 사람이 정의한 테스크에 한정되어 있습니다. 새로운 서브 테스크에서도 통할지 확인해야 합니다. 또한, 테스크 간의 결합으로 새로운 테스크가 나올 수도 있습니다.

Space Regularity

우리는 테스크를 샘플링하여 결과에 이용했는데, 더 넓은 공간에서의 샘플링을 했을 때도 결과가 그대로 나올지 확인해야 합니다.

Transferring to Non-visual and Robotic Tasks

모든 실험은 이미지에서만 이루어졌고 다른 시각적이지 않은 분야에서도 통할지 확인해야 합니다.

Lifelong Learning

계속해서 발전해나가고 있는 상황에서 Tasknomy가 언제까지 통할지 생각해야 합니다.