

# Transfer Learning tra Task Visuali – Progetto di Ricerca e Piano di Attività

Il riconoscimento di oggetti, la stima della profondità, il rilevamento dei bordi sono alcuni esempi di *task* visuali ritenuti importanti nella letteratura delle computer vision e classicamente affrontati dalla comunità dei ricercatori in maniera totalmente disgiunta. I principali *task* visuali, tuttavia, sono in forte collegamento tra loro. Per esempio, la normale delle superfici sono la derivata della profondità, la geometria di una scena è strettamente collegata alla tipologia di oggetti che sono contenuti in essa, i bordi sono correlati alla struttura 3D della scena. Negli ultimissimi anni, sono state sviluppate nuove tecniche di computer vision, e in particolare le reti neurali convoluzionali o CNN, in grado di risolvere in maniera particolarmente efficace tutti i principali *task* visuali. Nondimeno, i sistemi allo stato dell'arte continuano a affrontare i diversi *task* in maniera indipendente l'uno dall'altro, ovvero ignorando ogni possibile correlazione tra i vari *task*, con la conseguente necessità di grandissime quantità di dati etichettati manualmente per l'addestramento delle CNN. Invece, essere in grado di trasferire conoscenza fra reti orientate alla soluzione di *task* differenti potrebbe permettere di ottenere modelli basati su CNN in grado di ottenere ottime prestazioni a partire da quantità ben più limitate di dati di training, con un ovvio conseguente beneficio in termini, di tempo, risorse computazionali ed energia. Ciò nonostante, l'unico approccio di *task transfer* oggi ampiamente adottato dalla comunità scientifica consiste nel pre-allenare la CNN orientata a un dato *task target* sul dataset ImageNet<sup>1</sup> e per il *task* della classificazione di immagini. Sebbene questo approccio sia notoriamente utile, nessuno garantisce che sia il migliore, e ci potrebbero essere *task* molto più adeguati da cui trasferire la conoscenza.

Comprendere le relazioni tra i vari *task* visuali è quindi un campo di ricerca aperto, e ci sono ancora moltissime domande senza risposta. Un primo passo fondamentale in questa direzione è stato fatto da *Taskonomy*<sup>2</sup>, un lavoro davvero seminale, premiato con il Best Paper Award alla conferenza CVPR 2018, che ha proposto un approccio computazionale per creare un grafo di dipendenze tra *task* visuali. Questo grafo rappresenta quali *task* possono essere utili ad altri, ed in quale misura. Tuttavia, questo grafo è calcolato con l'assunzione di avere un unico dominio (o *dataset*) con tutti i *task* disponibili, e un modello di CNN fissato a priori. In scenari reali, invece, si dispone di tanti dataset differenti con solo alcuni *task* fruibili in un certo dataset. In questo scenario, sarebbe utile sapere non solo da quale *task*, ma anche da quale dominio trasferire la conoscenza. Ancora più auspicabile sarebbe di essere in grado di combinare le conoscenze dei vari *task* e domini per ottenere prestazioni ancora migliori. In ultimo, sarebbe interessante esplorare come cambia la relazione tra i *task* al variare dell'architettura neurale. In un successivo

---

<sup>1</sup> Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

<sup>2</sup> Zamir, A. R., Sax, A., Shen, W., Guibas, L. J., Malik, J., & Savarese, S. (2018). Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3712-3722)

lavoro, denominato AT/DT<sup>3</sup> e svolto presso il CVLab del DISI, si è iniziata ad indagare la possibilità di apprendere la funzione di trasferimento della conoscenza fra due *task* visuali con un'altra rete neurale, e come usare tale rete per trasferire conoscenza tra *task* al variare del dominio allo scopo di diminuire sensibilmente la necessità di dati annotati. Sebbene questo lavoro cominci ad esplorare la possibilità di apprendere la relazione tra due *task* al variare del dominio, esso è specificamente orientato al trasferimento di conoscenza tra due *task* noti, e non allo studio generico dello spazio dei *task*.

In questo progetto si propone quindi di approfondire gli studi sul *task transfer learning*. Un primo stadio verte sull'analisi delle dipendenze tra i *task* visuali senza l'assunzione di operare all'interno di un singolo dominio. Dato che alcuni *task* potrebbero essere collegati in modo differente in base ai *dataset* in analisi, il primo passo dello studio è mirato a ottenere un grafo, ovvero una *Taskonomy*, che varia in base ai domini in presi in considerazione. Questo risultato potrebbe essere ottenuto individuando delle rappresentazioni dei domini, con la quale pesare differenzialmente le varie connessioni tra i *task*, permettendo così di avere un grafo dinamico che si adatta in base ai *dataset* a disposizione. Successivamente, si studierà la possibilità di combinare le informazioni provenienti da *task* e *dataset* multipli, in modo da ottenere la miglior combinazione di conoscenza da trasferire. Nell'ultima fase del progetto sarà invece investigata la dipendenza del grafo dall'architettura scelta, con lo scopo di individuare - o apprendere mediante tecniche di AutoML, le eventuali architetture che permettano una migliore trasferibilità della conoscenza tra *task* differenti e correlati, o se esistano architetture leggere e efficienti che possano ottenere prestazioni simili a quelle di modelli onerosi dal punto di vista computazionale grazie al *task transfer learning*. Obiettivo ultimo del progetto è quindi ottenere una più profonda comprensione delle connessioni tra i vari *task* visuali così da poter stabilire automaticamente quale sia la migliore conoscenza da trasferire dati un insieme di *task* e *dataset*, nonché quale sia l'architettura neurale più adatta a tale trasferimento. Questa ricerca ha la potenzialità di avere un notevole impatto sulla comunità del *deep learning*, permettendo di diminuire sia il bisogno di grandi quantità di dati etichettati manualmente labellati sia i costi computazionali causati dal continuo allenamento di nuovi modelli neurali da zero.

---

<sup>3</sup> Ramirez, P. Z., Tonioni, A., Salti, S., & Stefano, L. D. (2019). Learning across tasks and domains. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 8110-8119).