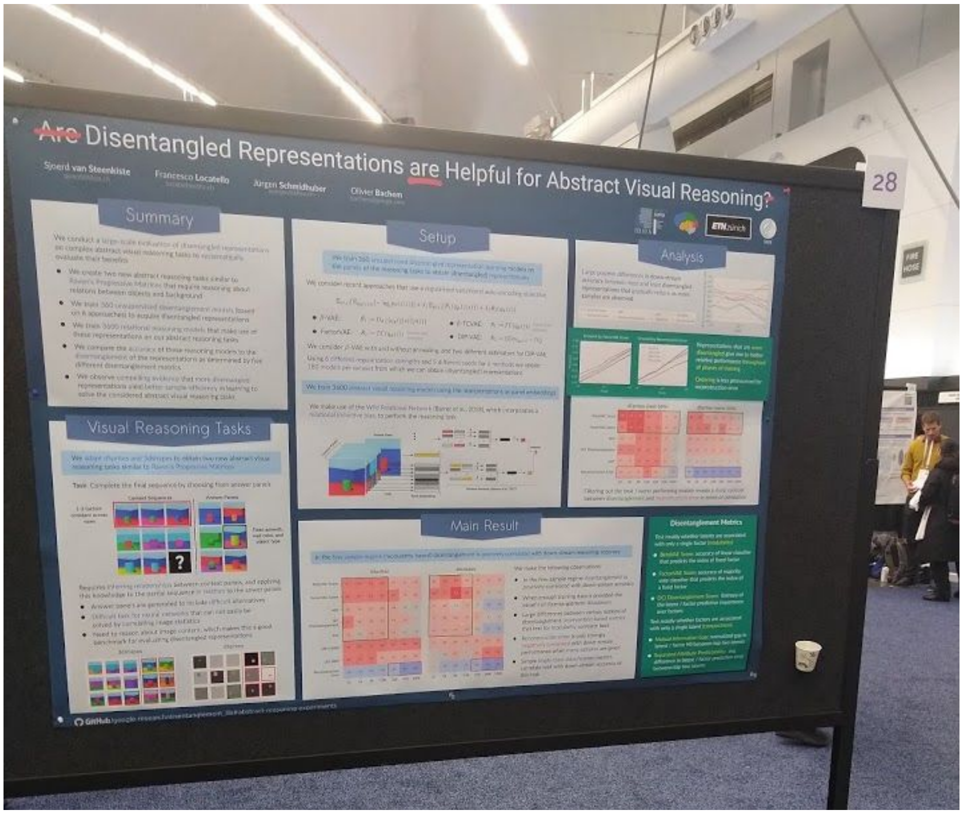
Disentangling disentanglement: Footnotes from NEURIPS - 2019

TL;DR : Disentangling disentanglement. Via this blogpost, I intend to try and summarize all of the dozen papers presented on disentanglement in deep learning in this year's NEURIPS.
Companion github repo replete with paper summaries and cheat sheets.
https://github.com/vinayprabhu/Disentanglement_NEURIPS_2019/

Background: Disentanglement in Representation learning

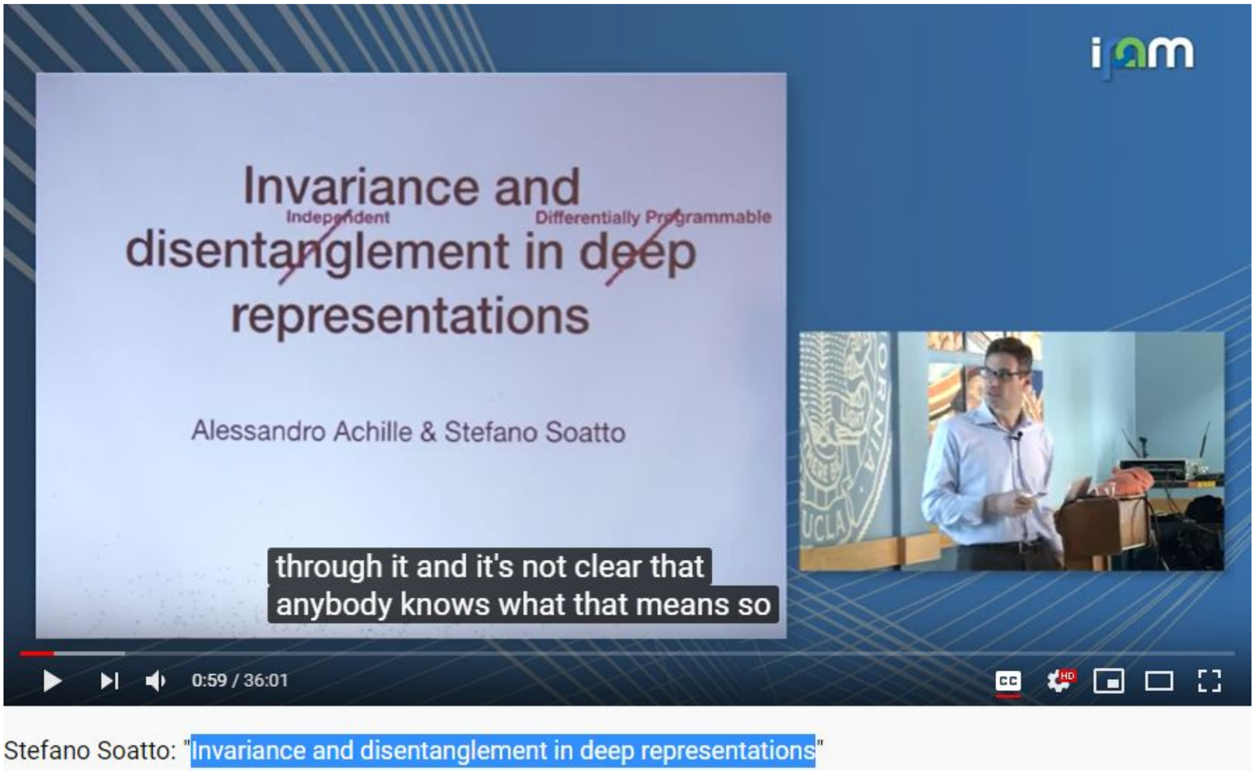
On Thursday evening, as I sauntered around the poster session in the massive East exhibition halls of the Vancouver convention center, I realized that I had chanced upon probably the 5th poster in the past couple of days entailing analysis of a disentanglement framework the authors had worked on. A quick check in the proceedings led me to this stunning statistic: A total of I-KID-YOU-NOT **dozen** papers were accepted this year with the term 'DISENTANGLEMENT' in the title. There were at least a few more that I chanced upon in the multitude of workshops. (There were 20+ papers and talks during the 2017 NEURIPS workshop on Learning Disentangled Representations: from Perception to Control - https://sites.google.com/view/disentanglenips2017)

I had first encountered this flavor of usage of the term in statistical learning in the last stages of my doctoral journey at CMU (circa 2013) when I read 'Deep Learning of Representations: Looking Forward' by Yoshua Bengio in which he emphasized the need to be '.. learning to *disentangle* the factors of variation underlying the observed data'.(How I wish he still authored such single author papers) [Link: https://arxiv.org/pdf/1305.0445.pdf]



As it turns out, much to the chagrin of the physicists perhaps, if you are working on teasing out visual style from digit type on MNIST, or separating *shape and pose in images of human bodies and facial features from facial shape on CelebA* or grappling with unwrapping the effects of mixture ratio of the two constituent compounds and environmental factors such as thermal fluctuation in images generated for microstructure growth, you are *disentangling.*

There seems to be no consensus on what the term precisely means or what metric(s) capture the extent of it, an observation that is confirmed by this rather funny/snarky slide in Stafano Soatto's talk at IPAM (refer to the playlist below)



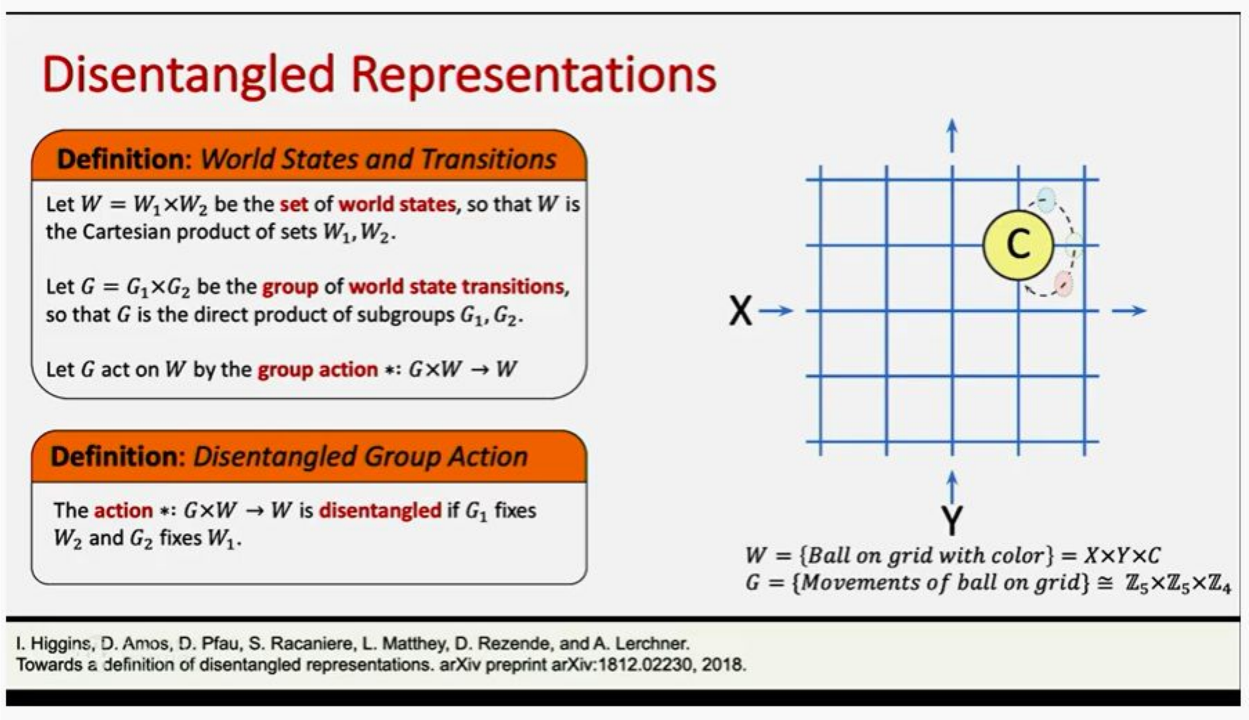
Invariance and
Independent Differentially Programmable
~~disentanglement~~ in ~~deep~~
representations

Alessandro Achille & Stefano Soatto

through it and it's not clear that
anybody knows what that means so

Stefano Soatto: "Invariance and disentanglement in deep representations"

That said, this is not a case of there existing a mere smattering of empirical experiments that all use their own customized notion of disentanglement. In fact, reasonably rigorous frameworks have been proposed harnessing powerful tools from areas such as Variational inference, Shannonian Information theory, Group theory and matrix factorization. Deepmind's group theoretic treatment of the same seems to have perched itself as one of the go-to frameworks. In case you are looking for a succinct 3 min recap of what this is, please refer to this video that I saw during one of Simons Institute workshops (around the 7th minute) [Video link:
https://www.youtube.com/watch?v=PeZIo0Q_GwE&t=1360s].
A much detailed talk from one the main authors (Dr.I.Higgins- Deepmind can be found here:
https://www.youtube.com/watch?v=XNGo9xqpgMo)



Disentangled Representations

Definition: *World States and Transitions*

Let $W = W_1 \times W_2$ be the **set** of **world states**, so that W is the Cartesian product of sets W_1, W_2.

Let $G = G_1 \times G_2$ be the **group** of **world state transitions**, so that G is the direct product of subgroups G_1, G_2.

Let G act on W by the **group action** $*: G \times W \rightarrow W$

Definition: *Disentangled Group Action*

The **action** $*: G \times W \rightarrow W$ is **disentangled** if G_1 fixes W_2 and G_2 fixes W_1.

$W = \{Ball\ on\ grid\ with\ color\} = X \times Y \times C$

$G = \{Movements\ of\ ball\ on\ grid\} \cong \mathbb{Z}_5 \times \mathbb{Z}_5 \times \mathbb{Z}_4$

I. Higgins, D. Amos, D. Pfau, S. Racaniere, L. Matthey, D. Rezende, and A. Lerchner.
Towards a definition of disentangled representations. arXiv preprint arXiv:1812.02230, 2018.

A bird’s view of the papers presented

In Fig 3 below, is a bird's-eye view of the work presented. I roughly bucketized them into two subsections depending on whether the *main* perceived goal of the paper was to either analyze and/or critique the properties of a pre-existing framework or to harness one and apply the same to an interesting problem domain. Bear in mind that this is admittedly a rather simplistic categorization and this is not very instructive of whether the applications oriented papers did or did not critique and analyze the frameworks used or that the analysis/critique papers did not include real-world applications.

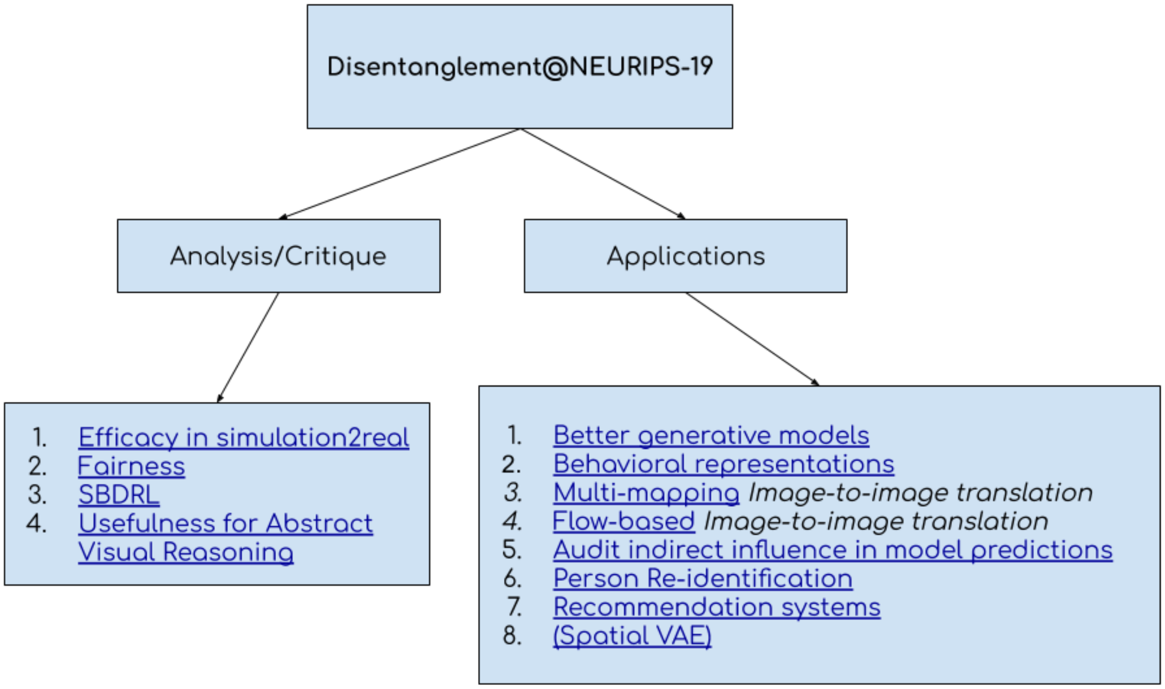

Disentanglement@NEURIPS-19
Analysis/Critique
Applications
1. Efficacy in simulation2real
2. Fairness
3. SBDRL
4. Usefulness for Abstract Visual Reasoning
1. Better generative models
2. Behavioral representations
3. Multi-mapping Image-to-image translation
4. Flow-based Image-to-image translation
5. Audit indirect influence in model predictions
6. Person Re-identification
7. Recommendation systems
8. (Spatial VAE)

Fig 3: Disentanglement papers categorization (NEURIPS -2019)

(You can find the pdf version with the paper links here: https://github.com/vinayprabhu/Disentanglement_NEURIPS_2019/blob/master/Disentanglement_papers_tree-diagram.pdf)

What do they mean by disentanglement?

In order to summarize the contexts in which disentanglement was used in these papers, I created a look-up-table (See Table-1). In those cases where the authors explicitly did not have a subsection dedicated to define the same, I improvised and extracted the gist (and hence the caveat [improv]).

Applications oriented	Disentanglement context
1. Explicit Disentanglement of Appearance and Perspective in Generative Models	Disentangled representation learning finds compact, independent and easy-to-interpret factors of the data
2. **Disentangled behavioural representations**	Disentanglement loss: The second term in the loss function favors disentangled representations by seeking to ensure that each dimension of the latent space corresponds to an independent factor of contribution in the variation across input sequences
3. Multi-mapping Image-to-Image Translation via Learning Disentanglement	[improv]We attempt to disentangle the images into solely independent parts: content and style. Moreover, we align these representations among image domains, which allows us to utilize rich content and style from different domains and manipulate the translation in finer detail.
4. Flow-based Image-to-Image Translation with Feature Disentanglement	Feature disentanglement: There are several origins of diversity within the image translations; some are from the given conditions (e.g., the mixture ratio of the two compounds) and the others are from environmental factors such as thermal fluctuation. For material property optimization, disentangling such diversity within those images is also important. [improv] We aim to disentangle the latent feature into (1) condition-dependent (referred to as “cross-conditional“ and “condition-specific“) and (2) condition-independent parts.
5. Disentangling Influence: Using disentangled representations to audit model predictions	The idea of a disentangled representation is to learn independent factors of variation that reflect the natural symmetries of a data set.
6. Learning Disentangled Representation for Robust Person Re-identification	The key idea behind the feature disentanglement is to distill identity-unrelated information from the identity-related feature, and vice versa.
7. Learning Disentangled Representations for Recommendation	Their framework posits that … user behavior data encapsulate latent user representations that merit both macro-disentanglement (ex: Product categories such as clothing) and micro-disentanglement (such as size and color)
8. Explicitly disentangling image content from translation and rotation with spatial-VAE	(Disentanglement is about) …finding data generative factors that encode semantic content independently from pose variables such as rotation and translation

Analysis / Critique oriented	Disentanglement context
1: Symmetry-Based Disentangled Representation Learning requires Interaction with Environments	[improv] Symmetry-Based Disentangled Representation Learning (SBDRL): Inspired by symmetry transformations in Physics that change only some properties of the underlying world state, while leaving all other properties invariant.
2: On the Fairness of Disentangled Representations	[Observations x are caused by k independent sources z1,...,zk] Disentanglement learning treats the generative mechanisms as latent variables and aims at finding a representation r(x) with independent components where a change in a dimension of z corresponds to a change in a dimension of r(x).
3: On the Transfer of Inductive Bias from Simulation to the Real World: a New Disentanglement Dataset	We assume a set of observations of a (potentially high dimensional) random variable X which is generated by K unobserved causes of variation (generative factors) (i.e., G → X) that do not cause each other. These latent factors represent elementary ingredients to the causal mechanism generating X. The most commonly accepted understanding of disentanglement is that each learned feature in Z (learnt latent embedding) should capture one factor of variation in G.
4: Are Disentangled Representations Helpful for Abstract Visual Reasoning?	A disentangled representation encodes information about the salient factors of variation in the data independently

Reproducibility and open-sourced code:

Given the strong growing trend towards open sourcing the code used to produce the results, 10 of the 12 author-groups shared their github repos as well. This is captured in Table-2 below:

Applications oriented	Code/ Open source
Explicit Disentanglement of Appearance and Perspective in Generative Models	https://github.com/SkafteNicki/unsuper/
Disentangled behavioural representations	N/A - Video: https://vimeo.com/368702071
Multi-mapping Image-to-Image Translation via Learning Disentanglement	https://github.com/Xiaoming-Yu/DMIT .
Flow-based Image-to-Image Translation with Feature Disentanglement	N/A
Disentangling Influence: Using disentangled representations to audit model predictions	https://github.com/charliemarx/disentangling-influence
Learning Disentangled Representation for Robust Person Re-identification	https://cvlab-yonsei.github.io/projects/ISGAN/
Learning Disentangled Representations for Recommendation	https://jianxinma.github.io/disentangle-recsys.html
Explicitly disentangling image content from translation and rotation with spatial-VAE	https://github.com/tbepler/spatial-VAE
Analysis / Critique oriented	**Disentanglement context**
1: Symmetry-Based Disentangled Representation Learning requires Interaction with Environments	https://github.com/Caselles/NeurIPS19-SBDRL
2: On the Fairness of Disentangled Representations	N/A
3: On the Transfer of Inductive Bias from Simulation to the Real World: a New Disentanglement Dataset	https://github.com/rr-learning/disentanglement_dataset
4: Are Disentangled Representations Helpful for Abstract Visual Reasoning?	https://git.io/JelEv

What now? Some ideas..

[Here are some scribbles to try and guilt myself into working on this more seriously. Please take these with a grain of salt or 12 :)]

1: Survey paper detailing the definitions, frameworks and metrics to be used.

2: Disentangling author / writing style / nation of origin using Kannada-MNIST dataset. (65 native volunteers from India and 10 non-native volunteers from USA)
https://github.com/vinayprabhu/Kannada_MNIST
3: It's somewhat surprising that no one's tried throwing a K user interference channel model for entanglement and see if an Interference Alignment [https://arxiv.org/pdf/0707.0323.pdf] like trick works for Dsprites-like datatsets
4: Disentangling Shoe type, pocket and device location from Gait representations
5: Bridging the body of work pertaining to (Hyperspectral) Unmixing / Blind source separation and disentangled representation learning.

Resource list:

Companion github repo replete with paper summaries and cheat sheets.
https://github.com/vinayprabhu/Disentanglement_NEURIPS_2019/
A.Datasets to get started with:

[1] https://www.github.com/cianeastwood/qedr
[2] https://github.com/deepmind/dsprites-dataset
[3] https://github.com/rr-learning/disentanglement_dataset

B. Video playlist:

[1] Y. Bengio's: From Deep Learning of Disentangled Representations to Higher-level Cognition
https://www.youtube.com/watch?v=Yr1mOzC93xs&t=355s
[2] \beta-VAE (Deepmind): https://www.youtube.com/watch?v=XNGo9xqpgMo
[3] Flexibly Fair Representation Learning by Disentanglement:
https://www.youtube.com/watch?v=nlilKO1AvVs&t=27s
[4] Disentangled Representation Learning GAN for Pose-Invariant Face Recognition:
https://www.youtube.com/watch?v=IjsBTZqCu-I
[5] Invariance and disentanglement in deep representations (Fun talk),
https://www.youtube.com/watch?v=zbg49SMP5kY

(From NEURIPS 2019 authors)
[1] The Audit Model Predictions paper: https://www.youtube.com/watch?v=PeZIo0Q_GwE
[2] Twiml interview of Olivier Bachem (3 papers on this topic at NEURIPS-19):
https://www.youtube.com/watch?v=Gd1nL3WKucY