How to make deep learning work in medical image registration: current advances, pitfalls and remaining challenges

Prof. Dr. Mattias Heinrich
- Medical Deep Learning Group
- Institute of Medical Informatics
- University of Lübeck, Germany

Thanks to my team in particular:
Lasse Hansen (PhD student geometric learning)
Hanna Siebert (PhD student DL registration)

Contact:
heinrich@imi.uni-luebeck.de
github.com/multimodallearning

Overview

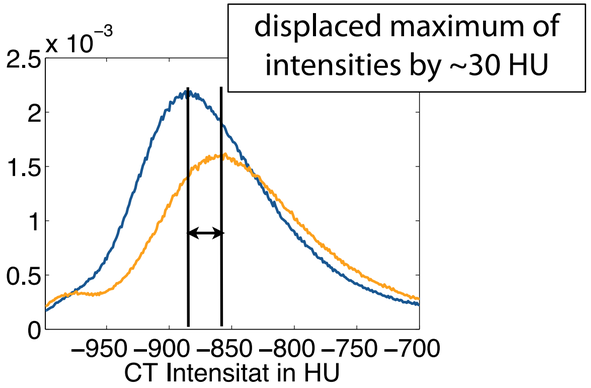
- **Motivation + Challenges**: large deformation, multi-modality (Learn2Reg)
- **Basic principles:** of DL registration in medical imaging
- **Advances / Pitfalls I**: differentiable resampling and instance optimisation
- **Advances / Pitfalls II**: synthetic supervision and label bias
- **Advances / Pitfalls III**: feature pyramids and two-stream architecture
- **Probabilistic Graph Networks** : discretise displacements / non-local loss, geometric networks using keypoints
- **Conclusions**

disclaimer: opinions based on personal experience

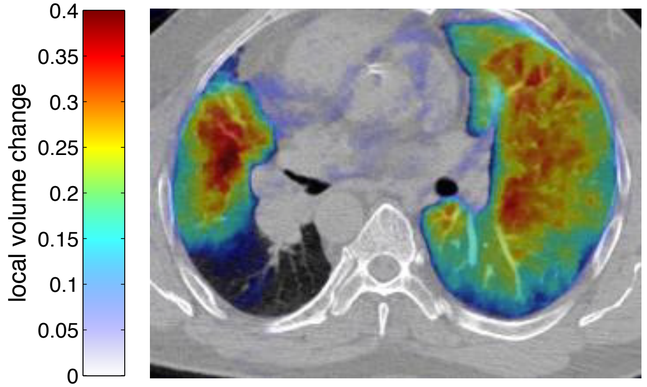
Motivation & Challenges

Measuring Temporal Change in Medical Images with Registration

Calculate local **lung ventilation** by **volume/intensity change of tissue** after inhale/exhale registration

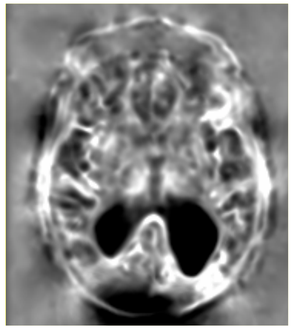


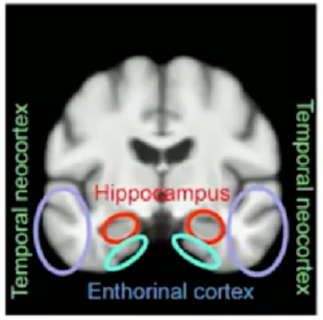
inhalation- / and expiration phase



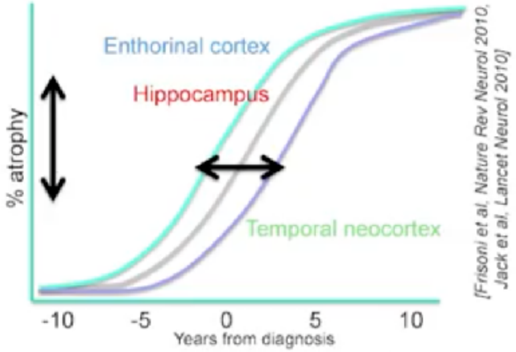
Heinrich et al.: "MRF-based Registration and Ventilation Estimation of Lung CT" IEEE Trans Medical Imaging 2013

Neurodegenerative disease analysis with deformable longitudinal registration





[Lorenzi 2011]



[Frisoni et al. Nature Rev Neurol 2010, Jack et al. Lancet Neurol 2010]

Determinant of Jacobi Matrix
second order gradients of deformation field **= volume change**
(brain atrophy in Alzheimer's disease, AD)

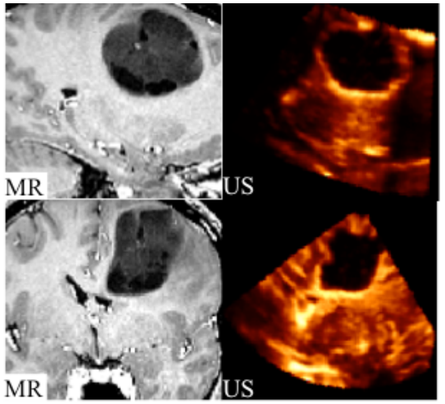
Temporal change of hippocampus in MRI scans is relevant biomarkers for AD detection *"average reduction compared to age-adjusted population is 12% hippocampus in MCI, and 24% in AD".**

→ segmentation of hippocampus (and its sub-parts) e.g. using U-Net

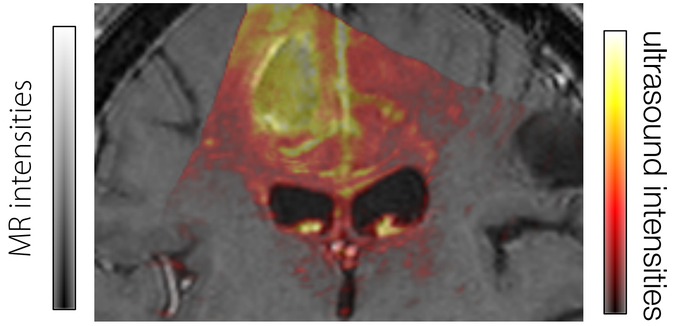
→ **determination of exact change over time using deformable image registration: subvoxel accurate and more robust and sensitive for small changes**

Combining multimodal information and modelling shape

→ **combine complementary strength of modalities** for image-guided interventions



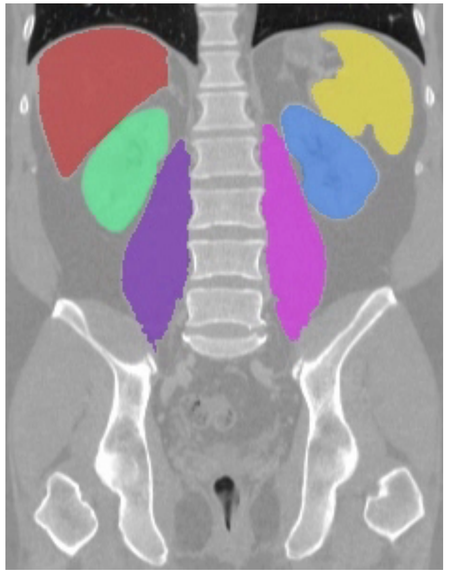
ultrasound-MR fusion for guided brain tumour surgery (MNI McGill)



overlay of inter-operative US over pre-treatment MR planning scan

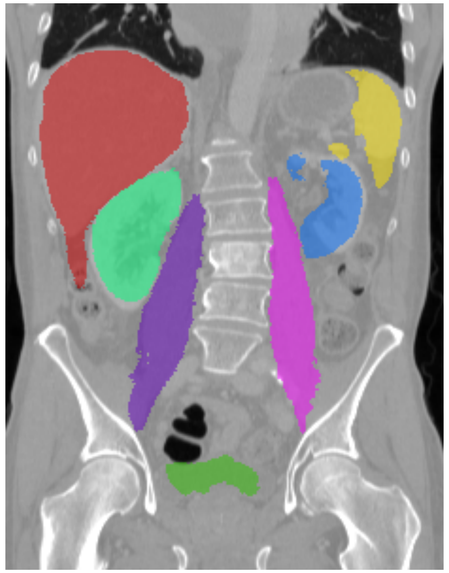
registration can implicitly use medical prior knowledge of **shape, spatial relations and appearance** by employing manually segmented atlases

atlas with expert segmentation



deformable registration between atlas and test scan

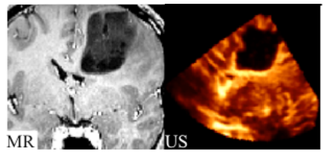
new test scan (abdominal CT)



transformation of expert segmentation to new scan

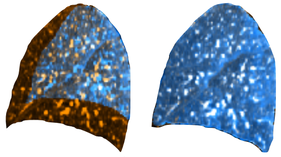
How much can medical 3D registration benefit from learning?

4 clinically relevant sub-tasks (datasets) complementary in nature and cover both intra- and inter-patient alignment, CT, ultrasound and MRI modalities, neuro-, thorax and abdominal anatomies



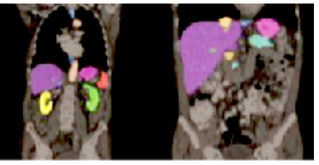
multimodal registration
large misalignment
CuRIOUS 2019 US-MRI registration challenge

Y Xiao, et al.: Evaluation of MRI to ultrasound registration methods for brain shift **TMI 2019**



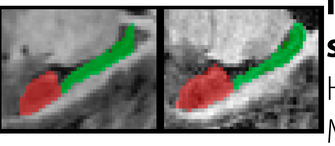
intra-patient lung motion inhale - exhale
relevant e.g. for COPD

A Hering, et al. Learn2Reg Challenge: CT Lung Registration - Training Data **2020**



inter-patient registration
large misalignment
Beyond the Cranial Vault MICCAI challenge

Z Xu, et al.: Evaluation of Six Registration Methods for Abdomen CT **TBME 2016**



large number of scans
subtle differences
Hippocampus MRI Medical Decathlon

AL Simpson, et al: "A large annotated medical image dataset for the development and evaluation of segmentation algorithms" **arXiv:1902.09063.**

difficulties for deep learning registration

1) noisy, limited or ambiguous labels
2) not one best network architecture established
3) compromise between generalisation, plausibility of transformations, accuracy and time has to be found

→ learning from small datasets
→ estimating large deformations
→ dealing with multi-modal scans / contrast variations

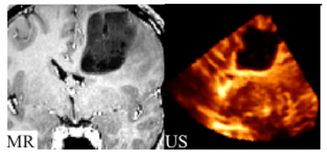
comprehensive and fair evaluation criteria

- **accurately and robustly transferring anatomical annotations / landmarks**
- **plausibility of deformations measured as transformation complexity** (the standard deviation of the local Jacobian determinant).
- **computation time:** due to differences in hardware, all steps of the employed pipeline are measured by running algorithms on the same CPU / Nvidia GPU backends (not a strict requirement for participants).

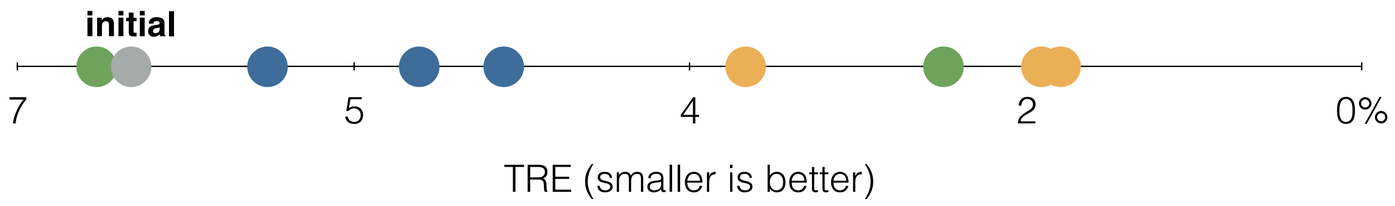
learn2reg.grand-challenge.org

Medical 3D Registration has yet to benefit from Deep Learning

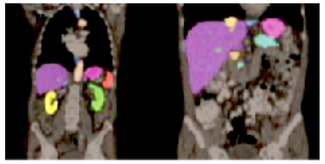
categorisation of methods: iterative (conventional)**, discrete** (conventional) **and deep learning**
→ clear advantage for **discrete conventional methods**, large gap for learning approaches



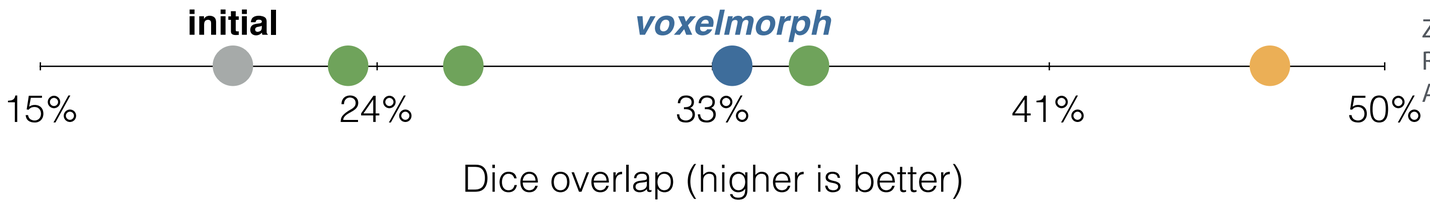
multimodal registration
large misalignment
CuRIOUS 2019 US-MRI registration challenge



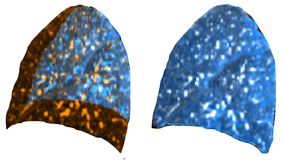
Y Xiao, et al.: Evaluation of MRI to ultrasound registration methods for brain shift **TMI 2019**



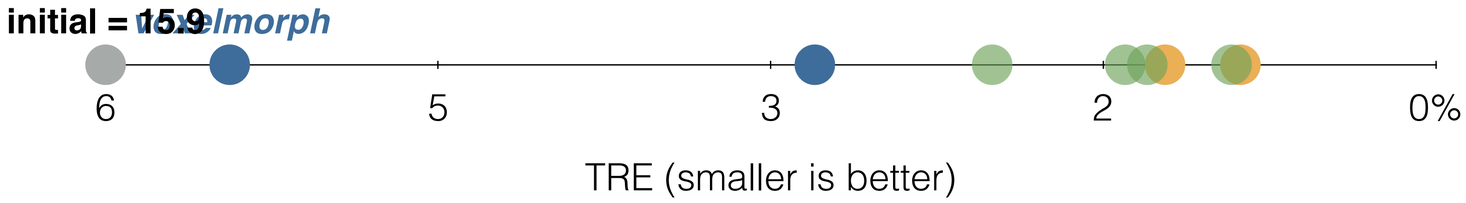
inter-patient registration
large misalignment
Beyond the Cranial Vault MICCAI challenge



Z Xu, et al.: Evaluation of Six Registration Methods for Abdomen CT **TBME 2016**



intra-patient lung
motion inhale - exhale
DIRLAB 4DCT + COPD



R Castillo, et al.: reference dataset registration spatial accuracy evaluation using COPDgene. **Phys Med Biol 2013**

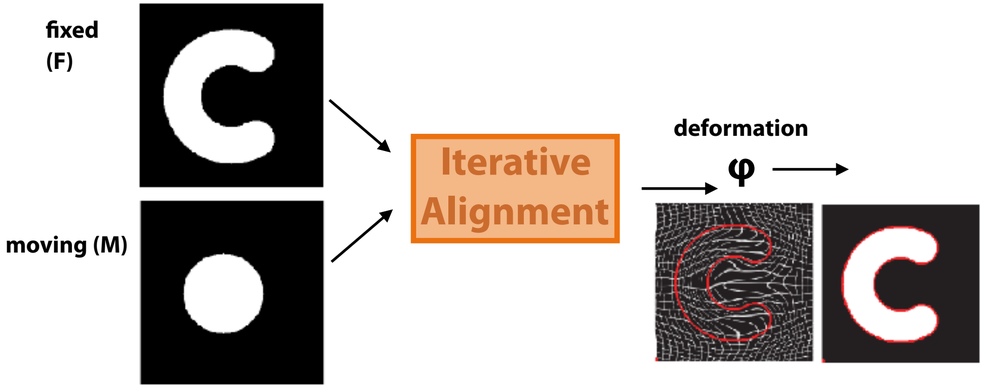
<u>difficulties for deep learning registration</u>

1) attempting to **train U-Net** for **features + optimisation** often **fails for large deformations**
2) **incorporating discrete displacement** search is **memory intensive** and backpropagation unfriendly
3) **recurrent DL registration suffers from vanishing gradients** and might be slow

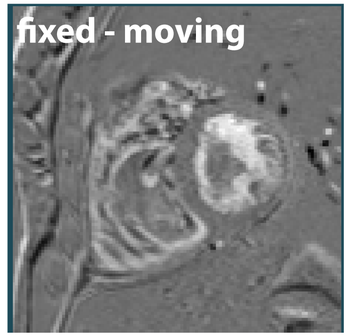
(MICCAI 2020): **learn2reg.grand-challenge.org**

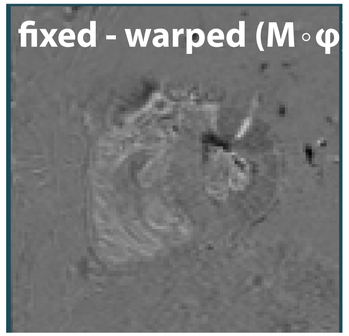
Basic Principles

Conventional Image Registration Iteratively Optimises Joint Cost



$$E(\varphi) = D(F, M \circ \varphi) + \lambda R(\varphi) \rightarrow \min$$





- **similarity metric** between fixed and moving image after we applied (nonlinear) transformation ***φ***
- ***mean-squared-error*** **(MSE)** is used for same modality
- **multi-modal fusion and improved edge alignment**: MIND, NGF, Mutual Information

given a pair of images: fixed (F) and moving (M)

1. we aim to find the best a **(nonlinear) transformation *φ***
2. that **maximises similarity** (minimises dissimilarity **D**) and
3. **penalises unrealistic deformations** (minimise regularisation **R**)

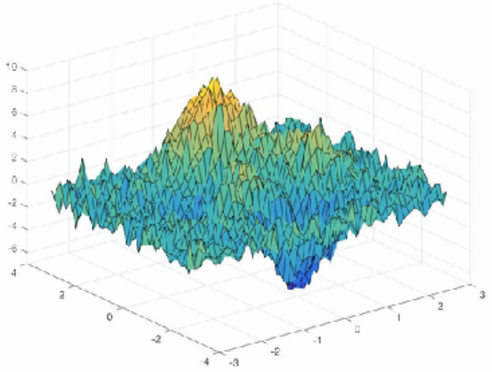
- image registration is inherently **ill-posed** (many plausible solutions) and there is usually **no** complete **ground truth**
- **many approaches are iterative** and use **multiple scales/resolutions**
- advanced methods with 2nd order **gradient descent** or **discrete optimisation**

even very efficient 3D medical registration tools require 30-60 secs. per registration

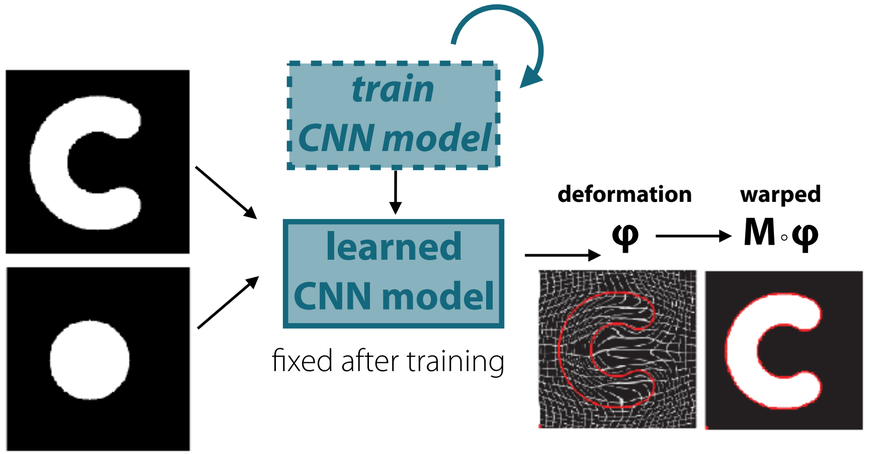
- ***diffusion regularisation (DR)*** is employed to penalise implausible transformations
- nonlinear displacement field ***φ* can contain irregularities** (folding of image content)
- S_x and S_y are x- and y-part of relative displacement field, DR penalises the **square of their gradients:**

$$R = \lambda \left(\frac{\delta S_x}{\delta x} \right)^2 + \lambda \left(\frac{\delta S_y}{\delta y} \right)^2$$

- finite differences: subtract pixels shifted by one



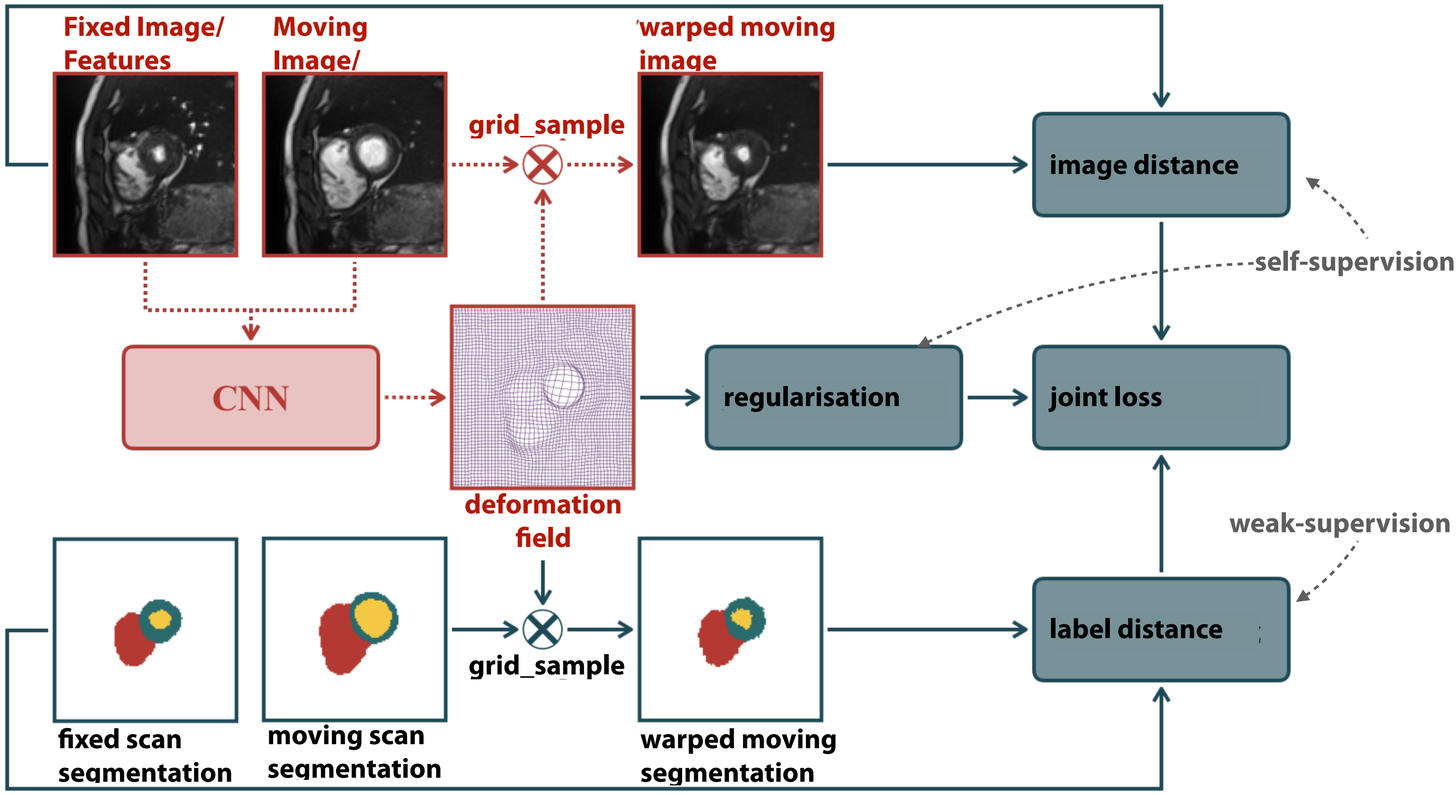
Principle of Learning an Optimisation CNN for Registration



- [deep learning registration](#) aims to **replace iterative steps with convolution layers**
- **time-consuming training** for CNN model
- **very fast in inference** (new data) ➞ msec.
- **generalise well** when trained with large datasets
- can be **improved with** expert **labels** (supervision)
- only work for image domain it was trained for

conventional methods work robustly across variety of tasks (depending on suitable cost functions, multi-level schemes + optimisation)

figure adapted from S Kuckertz, A Hering, MP Heinrich: "Enhancing Label-Driven Deep Deformable Image Registration .." ***BVM 2019***



green blocks (and expert labels) are only necessary during training,
red blocks remain fixed for inference

Advances / Pitfalls I:

differentiable resampling and instance optimisation

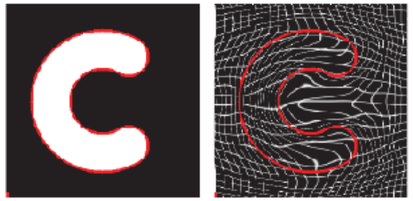
Gradients of displacements by differentiable sampling and metric

popularised in deep learning by Jaderberg's Spatial Transformer Networks

- conventional registration methods first **linearise the cost function** and **then apply iterative warping steps**
- it can be theoretically shown that **solving the functional directly using interpolation** is equivalent

The similarities in formulation and numerical realisation suggest a yet unexplained common background of warping methods on one hand, and methods with non-linearised constancy assumptions on the other hand. In fact we are able to demonstrate the equivalence of both schemes from the numerical point of view. If we restrict ourselves to spatial smoothness for the sake of simplicity, (8) reads calculated. Thus it is shown that the warping technique actually minimises an energy functional which is based on non-linearised constancy assumptions, whereas the minimisation relies on a coarse-to-fine strategy combined with two nested fixed point iterations.

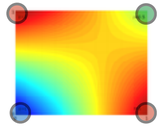
N. Papenberg, et al.: "Highly Accurate Optic Flow .. with Theoretically Justified Warping" ***IJCV 2006***



warped (M∘φ) deformation (φ)

- calculating (dis)-similarity metric requires off-grid interpolation of transformed moving scan
- used as loss in deep network, we need **gradient with respect** to each element of **deformation field** S_x, S_y
- **grid_sample uses bilinear interpolation** for 2D images (by default)

M Jaderberg, K Simonyan, A Zisserman: "Spatial transformer networks". ***NIPS* 2015**



output of individual element (off-grid) depends on 4 input grid points
→ bilinear interpolation has gradient

$$F(x,y) = (w_1 I(\lfloor x+S_x(i,j)\rfloor, \lfloor y+S_y(i,j)\rfloor) + w_2 I(\lceil x+S_x(i,j)\rceil, \lfloor y+S_y(i,j)\rfloor) + w_3 I(\lfloor x+S_x(i,j)\rfloor, \lceil y+S_y(i,j)\rceil) + w_4 I(\lceil x+S_x(i,j)\rceil, \lceil y+S_y(i,j)\rceil))$$

bilinear interpolation coefficients depend on S_x, S_y

$$w_1 = (\lfloor x+S_x(i,j)+1\rfloor - (x+S_x(i,j)))\,(\lfloor y+S_y(i,j)+1\rfloor - (y+S_y(i,j)))$$
$$w_2 = (x+S_x(i,j) - \lfloor x+S_x(i,j)\rfloor)\,(\lfloor y+S_y(i,j)+1\rfloor - (y+S_y(i,j)))$$
$$w_3 = (\lfloor x+S_x(i,j)+1\rfloor - (x+S_x(i,j)))\,(y+S_y(i,j) - \lfloor y+S_y(i,j)\rfloor)$$
$$w_4 = (x+S_x(i,j) - \lfloor x+S_x(i,j)\rfloor)\,(y+S_y(i,j) - \lfloor y+S_y(i,j)\rfloor)$$

floor/ceil are differentiable function
→ we can back propagate through grid sampling

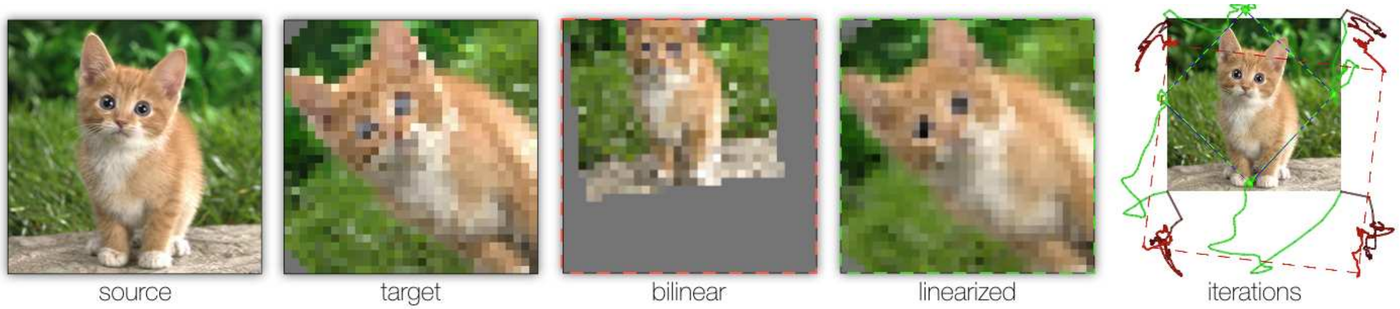
DL-Toolbox for Registration w/o Learning and Linearised Multi-Sampling

AIRLAB (autograd image registration lab)

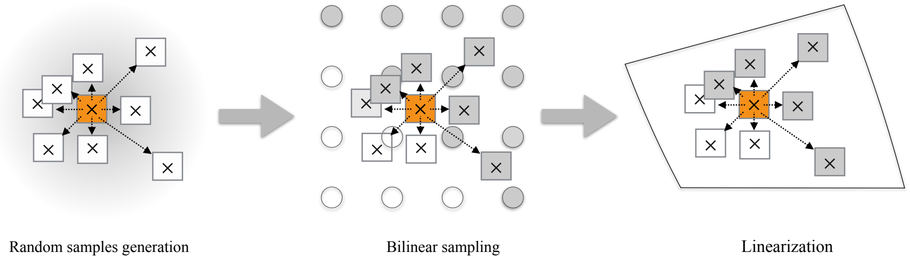
https://github.com/airlab-unibas/airlab
provides **comprehensive pytorch implementation of conventional registration algorithms**
→ GPU speed and ease of implementation
→ differentiable grid_sample enables optimisation of **arbitrary**

AIRLAB
Autograd Image Registration Laboratory

R Sandkuehler, et al: "AirLab: Autograd Image Registration Laboratory". **arXiv** 2018



bilinear sampling produces very sparse and noisy gradients (only 4 pixels involved)
transformed source image is shown at convergence (right): optimisation path (iterations) show: →
linearised sampling provides improved gradients and leads to better convergence



$$\mathbf{A}_i = \left(\mathbf{X}_i^\top \mathbf{X}_i + \epsilon \mathbf{E}\right)^{-1} \mathbf{X}_i^\top \mathbf{Y}_i$$

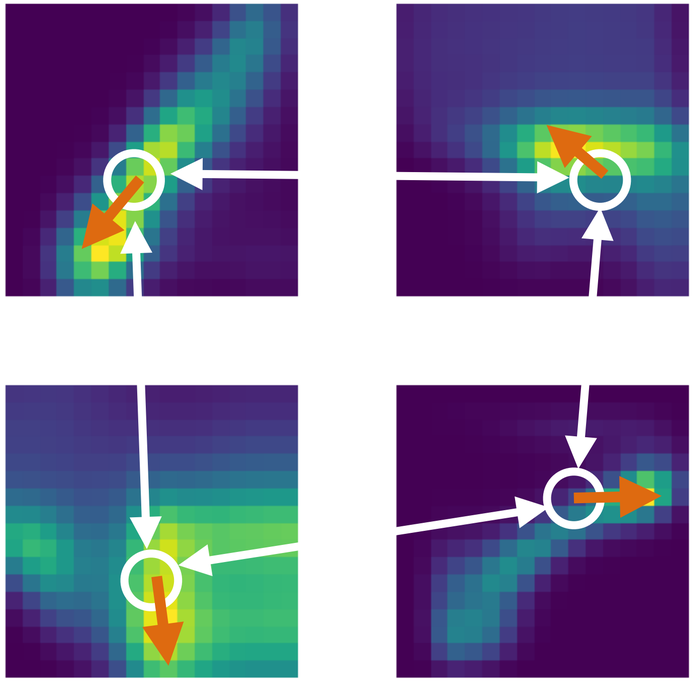
$\mathbf{X}_i$ coordinates 2D+1 ***around centre*** 8x3
$\mathbf{Y}_i$ ***differences*** of intensities at sample points 8x1
E identity matrix (Tikhonov reg.) 3x3
→ $\mathbf{A}_i$ ***local gradient*** of sampled intensity w.r.t. x

Linearised multi-sampling generates a set of **random auxiliary samples** for each pixel (computing intensities by bilinear interpolation) and **fits a 3D plane using least squares** (to robustly estimate slope)

W Jiang, et al.: "Linearized Multi-Sampling for Differentiable Image Transformation" **ICCV 2019**

Instance optimisation: use Adam for continuous displacements

<u>within 1 iteration:</u>
each spatial grid point
samples one value from 3D cost maps
→ vector gradient



$$\mathscr{C} \in H \times W \times D \times 15 \times 15 \times 15$$

pre-computed 6D cost tensors

$$\varphi^* := (x + \Delta x^*, y + \Delta y^*, z + \Delta z^*)$$

fine-tuned continuous displacement φ*
(initialised with softargmax)

joint energy functional
displacement cost + diffusion regularisation

$$L_{\text{instance}} = \mathscr{C}(x, y, z, \Delta x^*, \Delta y^*, \Delta z^*)$$

Key elements of instance optimisation:
1) softargmax for initialisation (requires discrete cost tensor)
2) **sample continuous displacement on pre-computed cost tensors**
3) 30-50 iterations of **Adam** with regularised energy

CRF smoothing (single forward path of regularisation network) **= equivalent to patch similarity and tricubic differentiable sampling** (see later slides)
→ more stable gradient than trilinear cf. STN & AirLab)

W Jiang, et al. "Linearized multi-sampling for differentiable image transformation." **CVPR 2019**

Sandkühler, R. et al. "AirLab: autograd image registration laboratory." **arXiv** 1806.09907 **2018**

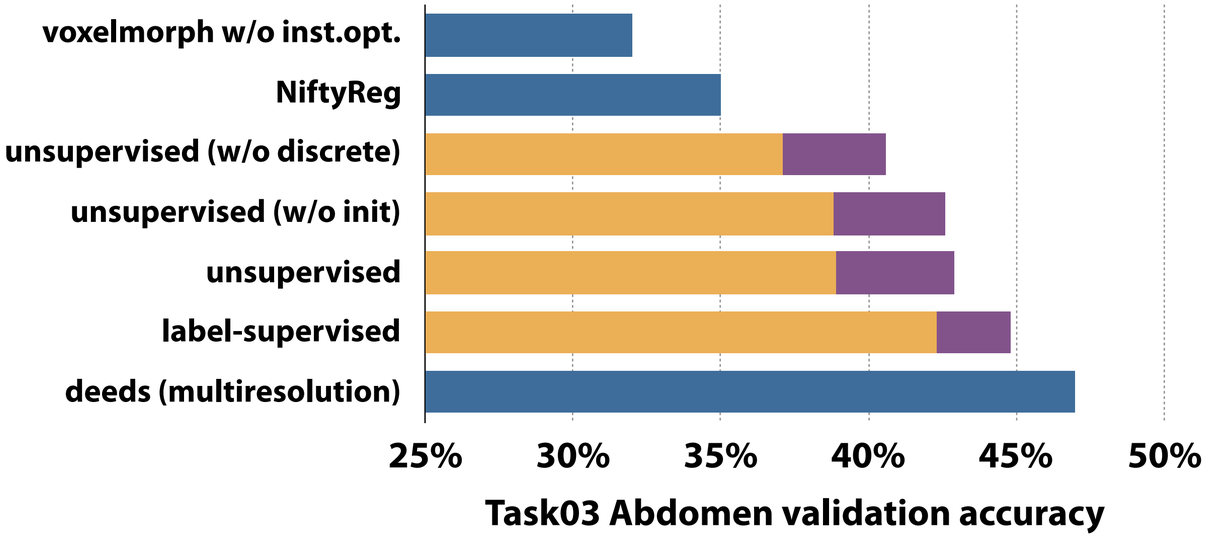
Benefits of instance optimisation on pre-computed cost tensors

Average Dice Task03 Abdomen validation after 1st and 2nd warp
compared to **Voxelmorph** and
conventional tools **NiftyReg + deeds**

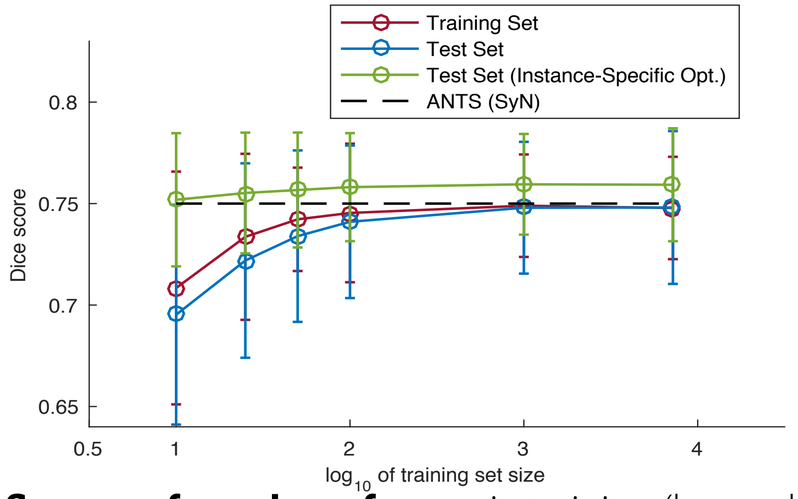
→ only 2% lower overlap **without label supervision**
→ **initialisation** of discrete optimum after CRF **negligible** (w/o init)
→ using **only continuous optimisation without** smoothing of **discrete displacement** dimensions we **outperform NiftyReg by 5% points**

very fast ~0.2sec since feature extraction is only performed once

+MP Heinrich, L Hansen: "Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5D displacement search", **MICCAI 2020**



Alternative in VoxelMorph: instance optimisation of U-Net parameters (green vs blue line)



influence of number of scans in training (log-scale)

- 1000 brain scans were required to reach accuracy **(blue line)** of conventional registration ANTS
- note: instance-specific requires iterative optimisation **(green line)** for new scans (slower than PDD because it requires backpropagation through U-Net)

G Balakrishnan, et al.: "VoxelMorph: A Learning Framework for Deformable Medical Image Registration" ***IEEE TMI 2019***

Advances / Pitfalls II:

synthetic supervision and label bias

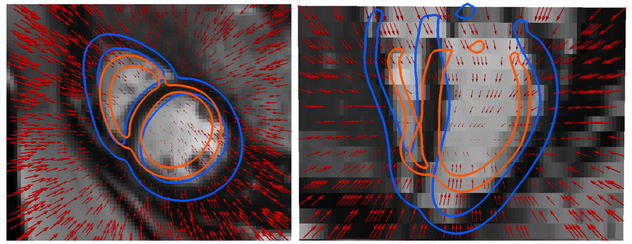
Examples of Registration Learning Synthetic Ground Truth

1) Computer Vision (2D Optical Flow)

- large networks require **huge number training scans** *with some ground truth*
- Sintel was striving for realism, FlyingChairs & Things3d generate >>1000 **arbitrary motion scenes** with **randomly (dis)placed 3D objects**
- use **random textures,** backgrounds + **post-processing** with simulated blurs, glares and contrasts

N. Mayer et al.: "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation" ***CVPR 2016***

2) Cardiac motion models using manual segmentation and surface matching

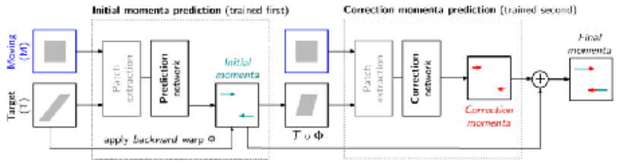


Rohe et al.: "SVF-Net: Learning Deformable Image Registration Using Shape Matching" ***MICCAI 2017***

training database with 187 patients, **pairwise generation of statistical motion fields**

3) Mimic a conventional method for speed

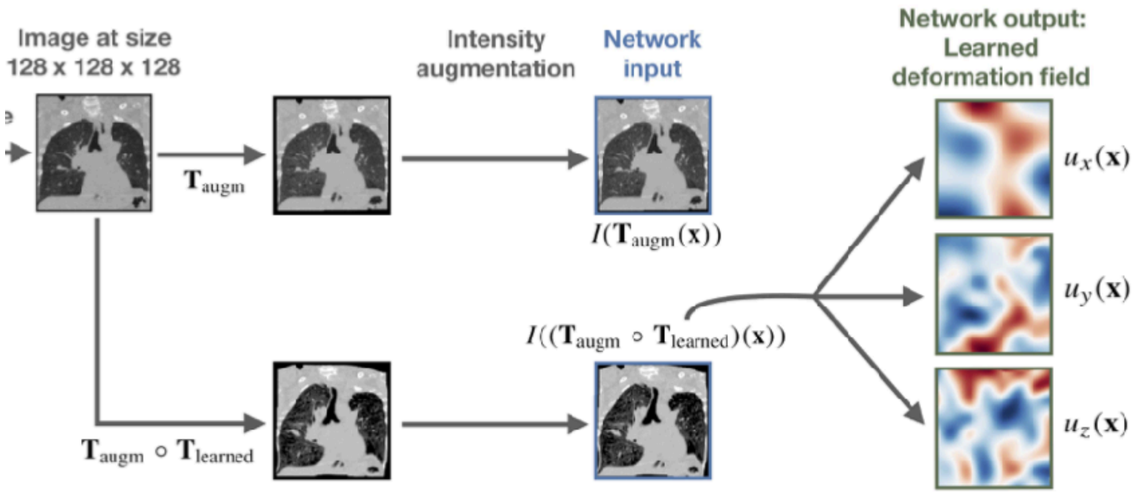
target and loss are defined by displacements of another algorithm e.g. LDDMM (not strictly synthetic ground truth)



X Yangt, .. M Niethammer: "Quicksilver: Fast Predictive Image Registration – a Deep Learning Approach" ***NeuroImage 2017***

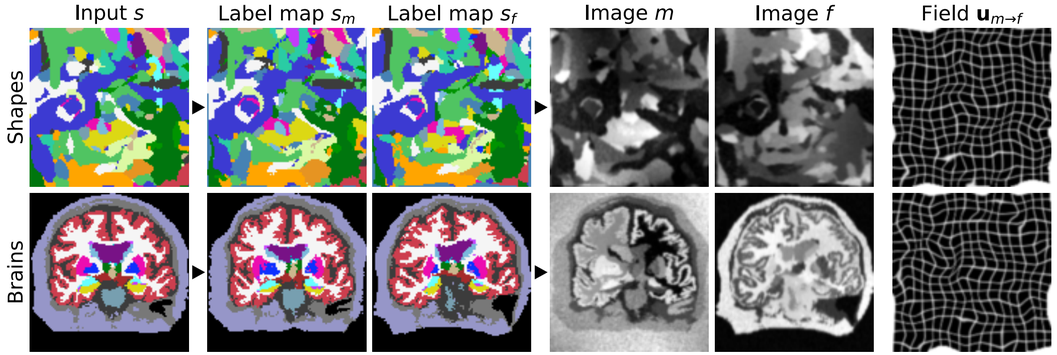
Examples of Registration Learning Synthetic Ground Truth

4) On-the-fly training set construction using intensity transforms and **random deformations as augmentation**
image is again deformed, using the learned transformation
→ *further alternative are GANs for deformation generation*

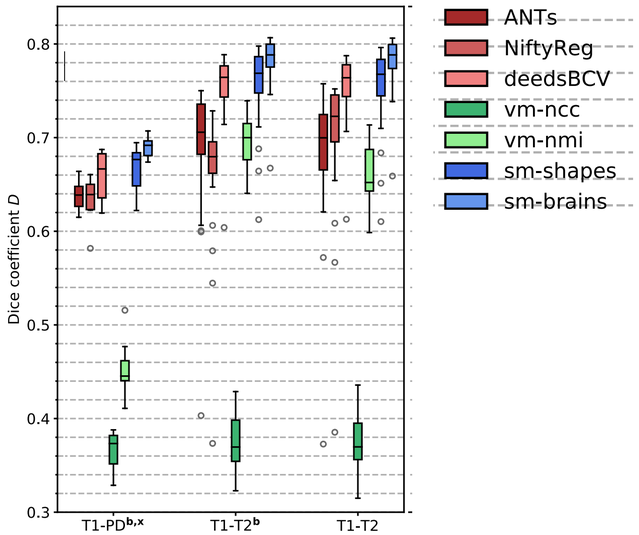


Eppenhof and Pluim : Pulmonary CT Registration Through Supervised Learning With Convolutional Neural Networks, ***TMI 2019***

5) Creating synthetic images (GANs or supervoxels)
create **multimodal images** with multiple smooth random noise channels
→ *labels are defined as argmax across these channels*



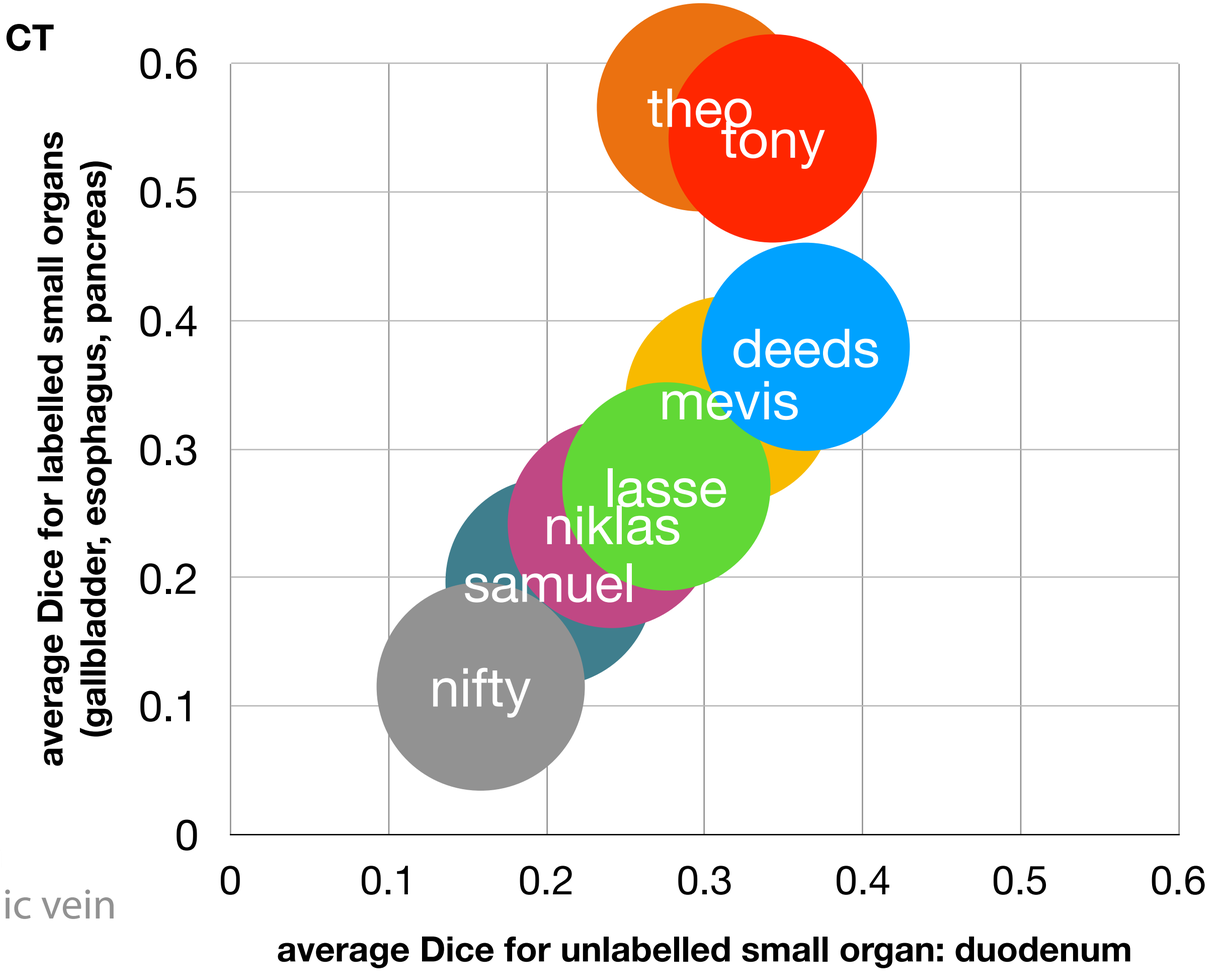
M. Hoffmann et al.: "Learning multimodal registration without real images"
https://**arxiv.org**/pdf/2004.10282

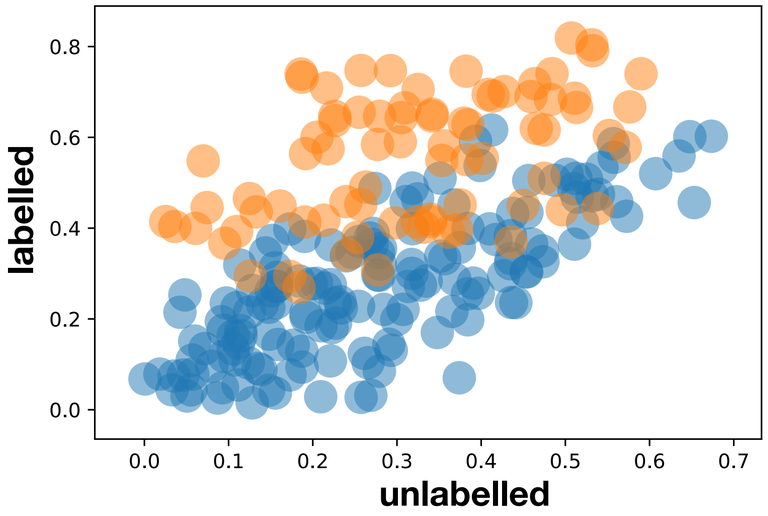


Label bias experiment (one more thing)

Task 03 Abdomen CT

1. spleen
2. right kidney*
3. left kidney
4. ***gallbladder***
5. ***esophagus***
6. liver
7. stomach
8. aorta
9. inferior vena cava
10. portal and splenic vein
11. ***pancreas***
12. right adrenal gland
13. left adrenal gland
14. **duodenum**

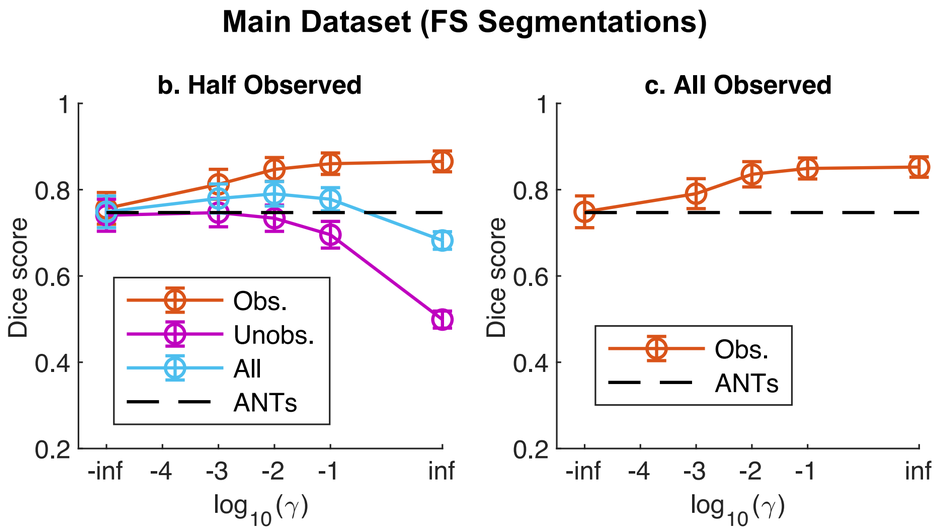


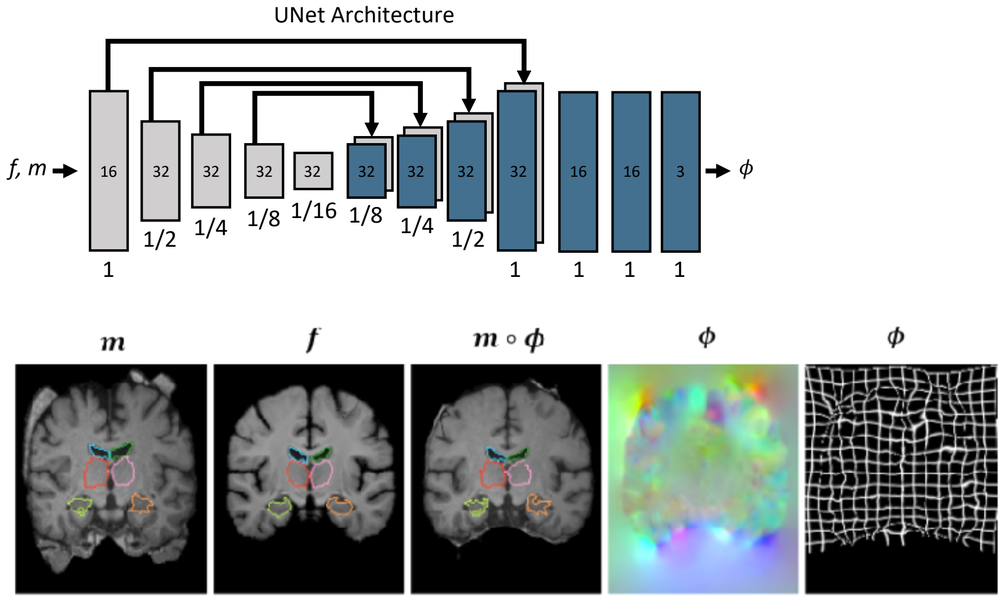


blue = unsupervised
yellow = label supervised

Label bias experiment (oberserved vs non-oberserved)

VoxelMorph (uses single-stream DL-reg architecture - with U-Net and NCC-metric plus label-loss)
Evaluation on large-scale brain registration study





influence of weak label-supervision

- when increasing the weight of label loss the **quality for other structures (unobserved) deteriorates**
  → supervision enables fine-tuning of some anatomies at the cost of others

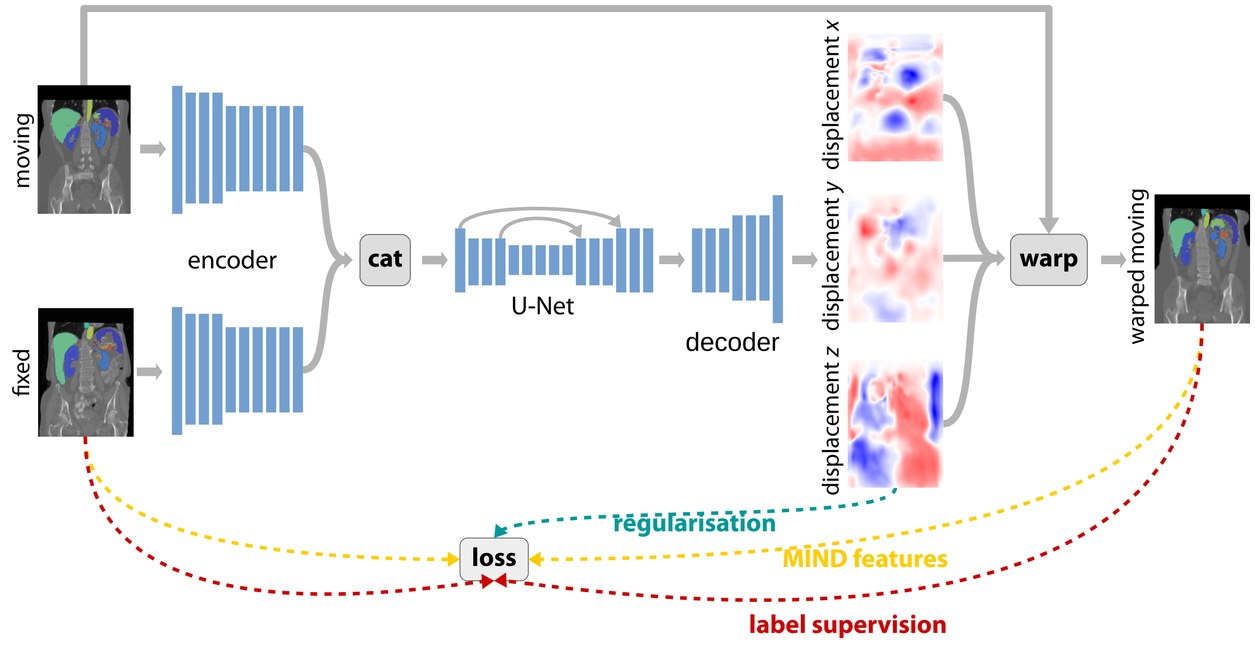
G Balakrishnan, et al.: "VoxelMorph: A Learning Framework for Deformable Medical Image Registration" ***IEEE TMI 2019***

further related work: Y. Hu, et al.: "Weakly-supervised convolutional neural networks for multimodal image registration" ***Med Imag Anal 2018***

Advances / Pitfalls III:

feature pyramids + two-stream architectures

Two-stream vs single-stream architecture



concatenating input images directly may complicate feature learning

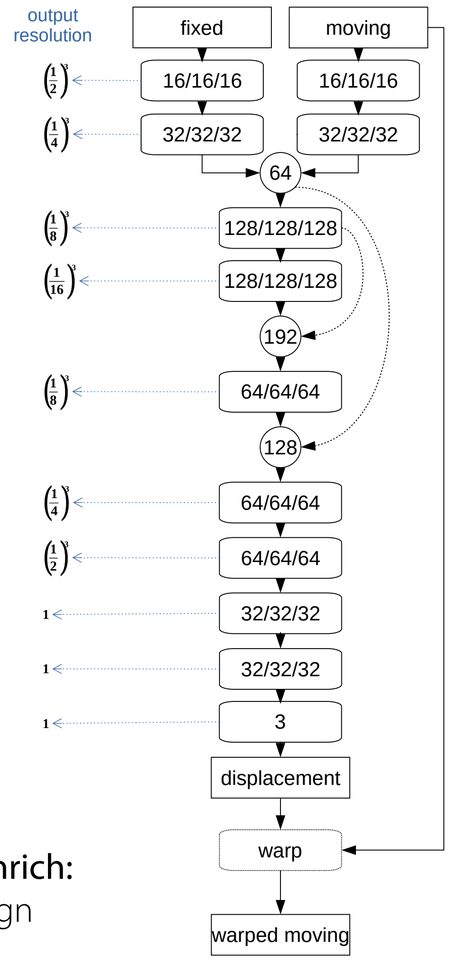
starting from affine alignment 25% Dice

- our compared unsupervised **single-stream architectures** (incl. VoxelMorph) achieve an average **increase** of up to **7% points**
- adding **two stream** leads to a **further increase of 4% points**
- **label supervision adds another 9-10% points**
  → **substantially more parameters** 4.2 million vs 400k of Voxelmorph

independent input encoder-blocks

U-Net

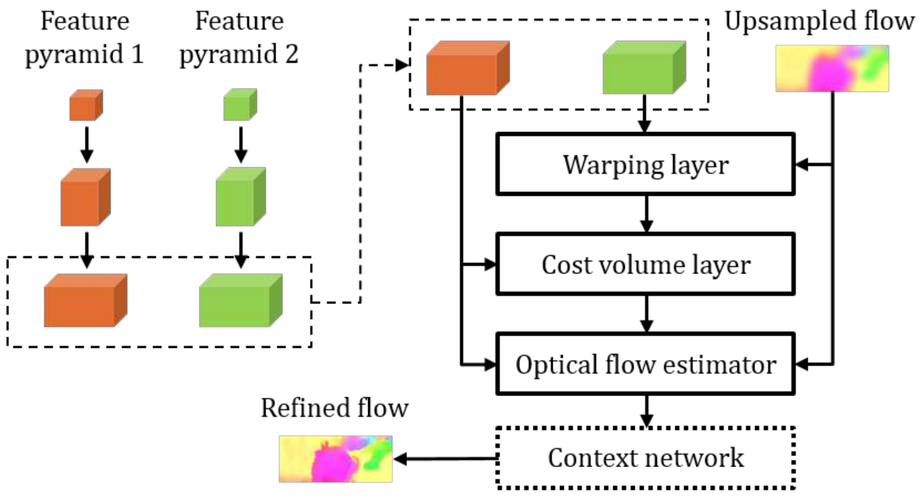
prediction decoder-blocks



Hanna Siebert, Lasse Hansen, M Heinrich: "Architecture matters: evaluating design choices for deep learning registration networks " 2020 ***under review***

two-stream details

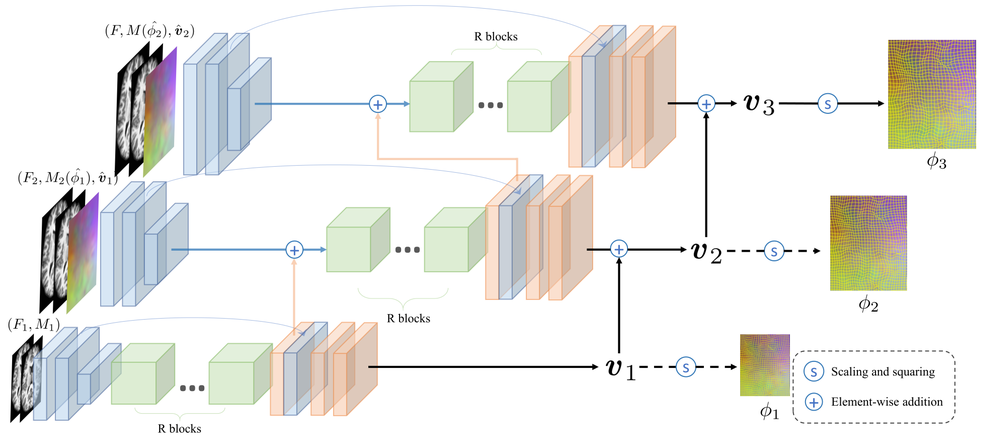
Feature Pyramids + multiple Warps



2D PWC-Net (outperformed FlowNet2 by large margin)
- at each pyramid level **features are warped with upsampled (previous) flow**
- a discretised **cost volume** (similar to correlation layer) is computed and processed using CNNs
- context network refines the (continuous valued) flow using large **receptive field dilated convs.**

D Sun et al.: "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume" ***CVPR 2018***

https://github.com/cwmok/LapIRN **Winner of Learn2Reg challenge**



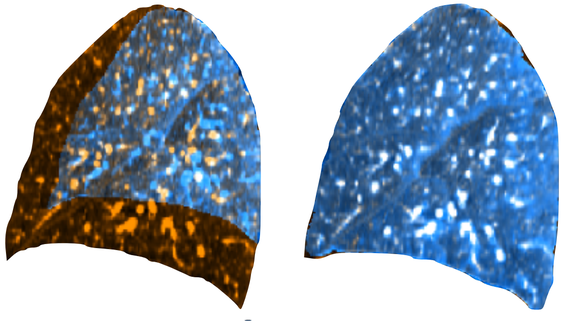
3D Lap-IRN (outperformed VoxelMorph by large margin)
- at each pyramid level **features are warped with upsampled (previous) flow**
- **supervision** is performed **using "metric" pyramid**
- each **ResNet blocks** contains **5 convolutions + lReLU** (followed by scaling-squaring per resolution)

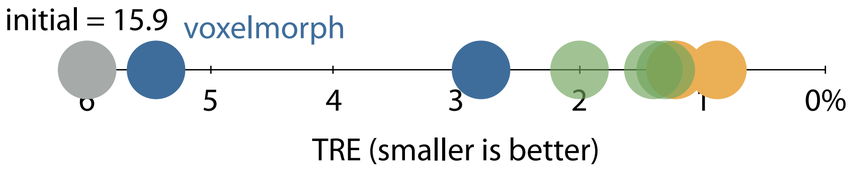
Tony Mok, Albert Chung: "Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks" ***MICCAI 2020***

Probabilistic Graph Networks : discretise displacements / non-local loss, geometric networks using keypoints

How to learn abdominal & lung registration?

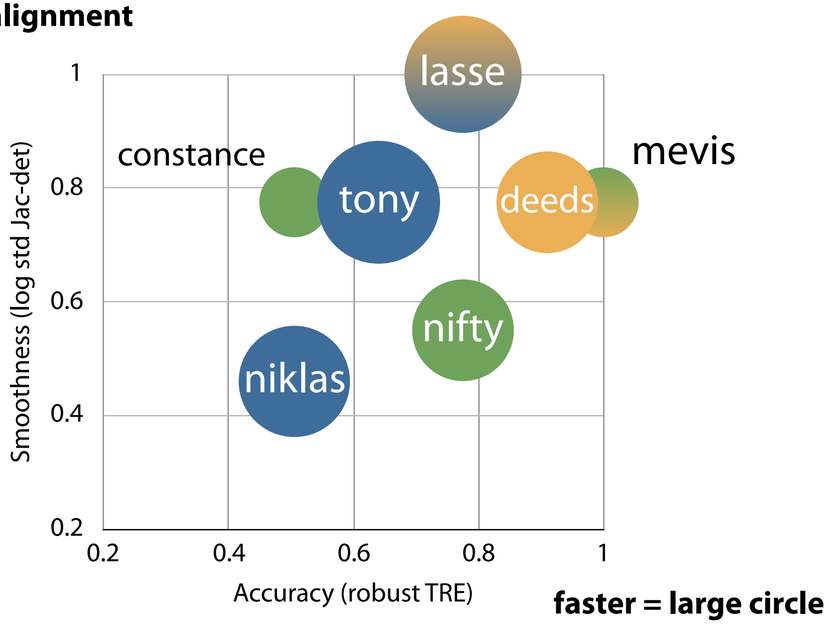
categorisation of methods: iterative (conventional)**, discrete** (graph-based) **and deep learning**
→ clear advantage for **discrete graph-based methods**, large gap for learning approaches





intra-patient lung motion
inhale - exhale
DIRLAB 4DCT + COPD

Learn2Reg Challenge Task 02 **inhale-exhale lung CT alignment**



faster = large circle

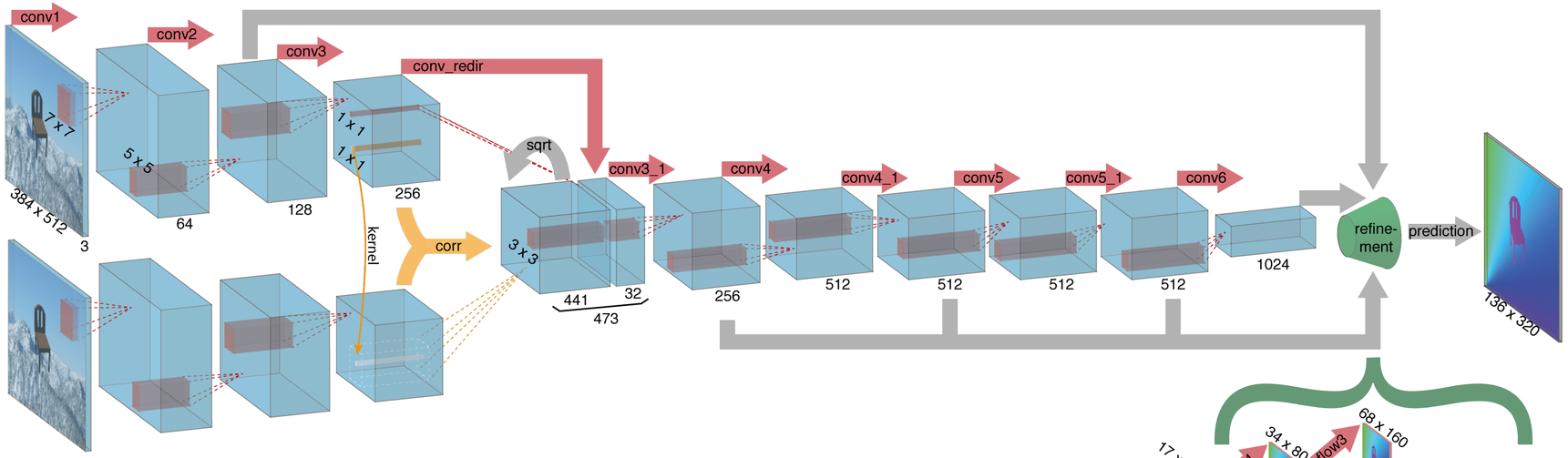
→ **our approach combines graphical models with learning**

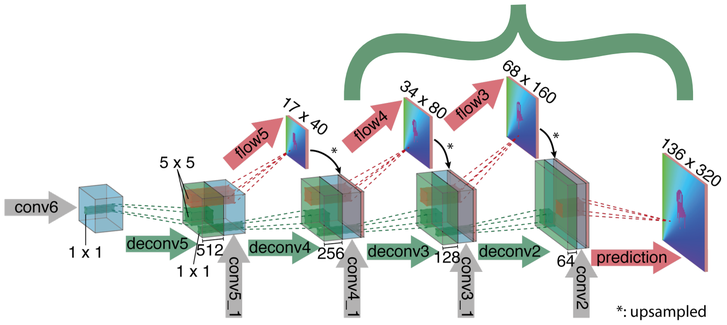
R Castillo, et al.: reference dataset registration spatial accuracy evaluation using COPDgene. **Phys Med Biol 2013**

Large deformations with discrete displacements

P Fischer, et al.: "FlowNet: Learning Optical Flow with Convolutional Networks" ***CVPR 2015***

FlowNetCorr





(large motion) correlation layer

- previous approaches are **limited in capture range**, by receptive field and limited number of conv. layers
- **FlowNetC** uses correlation layer without trainable weights but computation of (CC)-**metric over 441 discrete displacements** at once
- in addition FlowNet uses **deep supervision**, i.e. a loss at multi-resolution levels

trained on millions of synthetic image pairs (see earlier slides)

Disentangled feature learning for 3D discrete registration

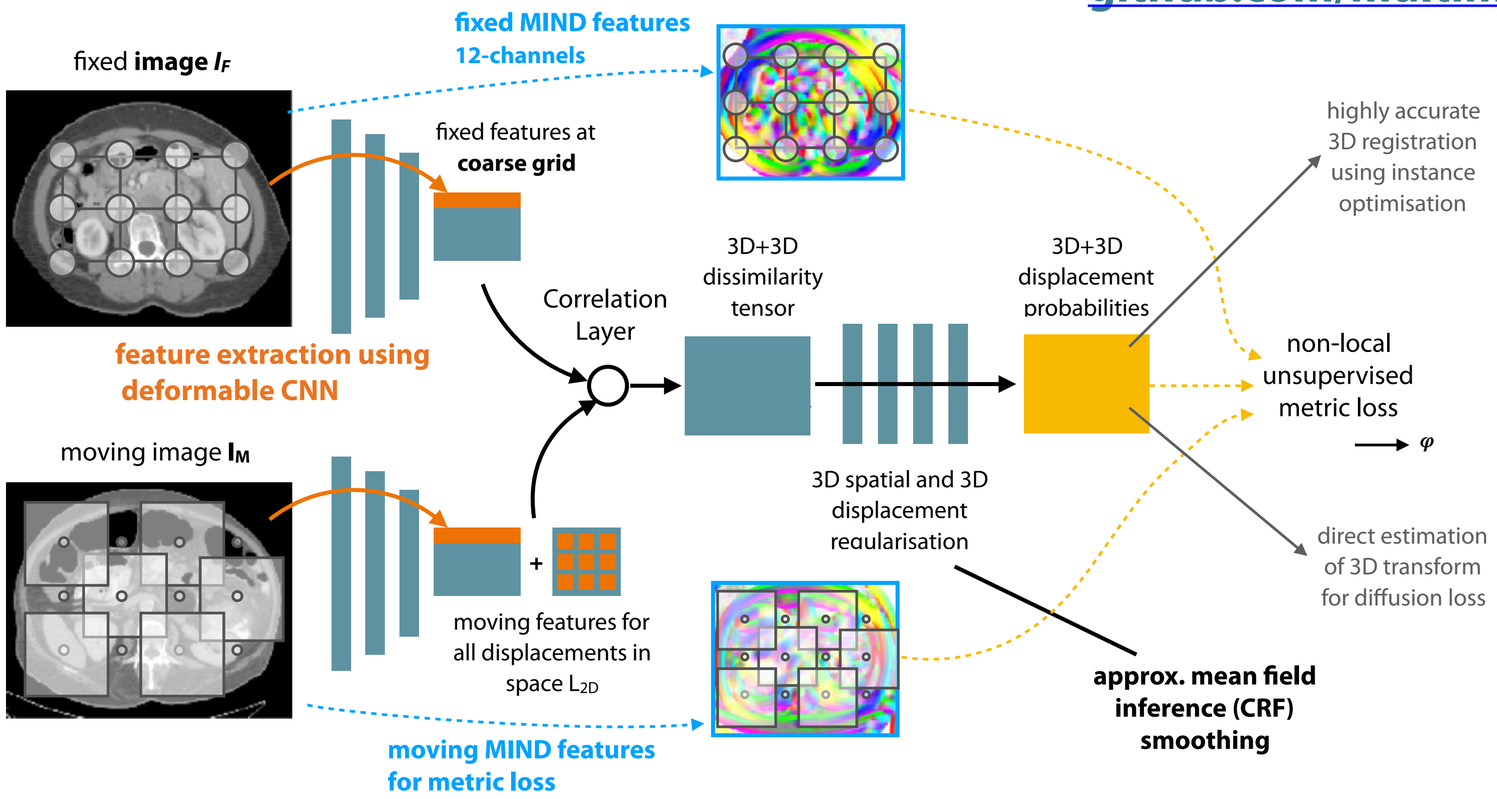
github.com/multimodallearning

probabilistic dense displacement (PDD)+ net

1) **OBELISK*** deformable convolutions for **feature** extraction
2) **Correlation Layer** discretised search space
3) **Spatial + Displacement regularisation** (smoothing) ~CRF optimisation

+MP Heinrich, L Hansen: "Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5D displacement search", **MICCAI 2020**

*MP Heinrich, O Oktay, N Bouteldja: "OBELISK-Net, Fewer Layers to Solve 3D Multi-organ Segmentation with Sparse Deformable Convolutions", **Medical Image Analysis**, 54, 1-9 **2019**



new solutions for highly accurate unsupervised DL registration

1) **discretised search space,** densely sampled 3D space **enables robust capture range**
2) further improvements using **iterative instance optimisation**

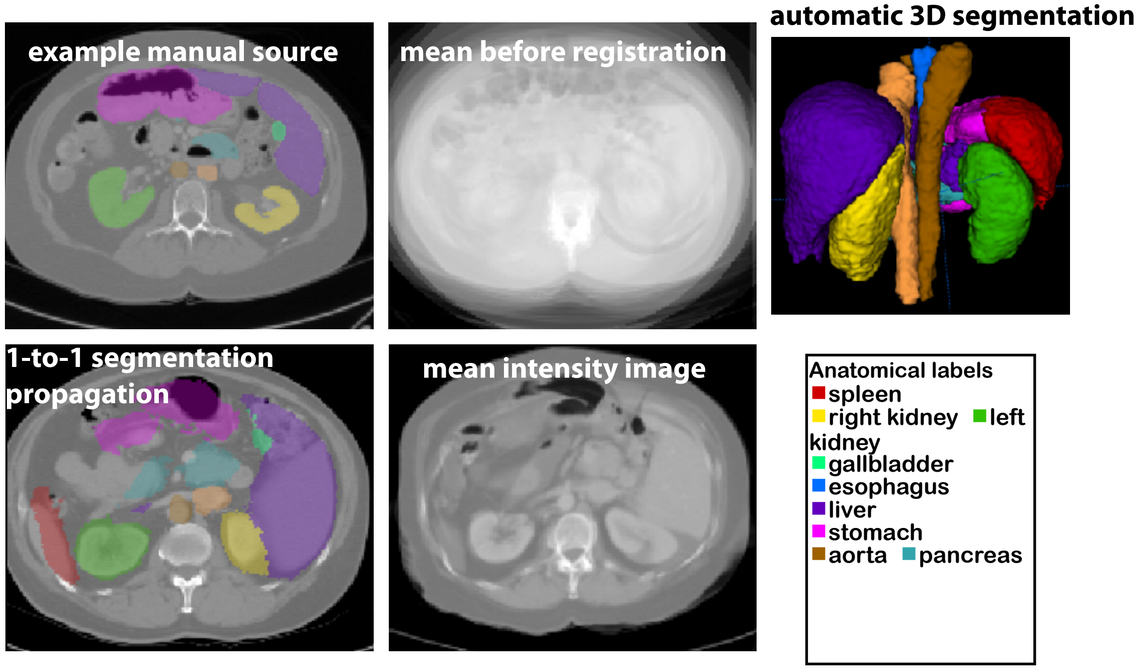
Developed for abdominal CT reg. - works across variety of tasks

one-to-one inter-patient alignment with very large initial misalignment and small anatomies for evaluation

10% points improvements compared to VoxelMorph and NiftyReg in particular for kidneys, stomach and pancreas

Method	avg(13)	HD95	detJ	memory	infer.	time
affine pre-reg.	28.1±8.3	14.6		train	GPU	CPU
Voxelmorph (MIND)	34.0±9.2	12.8	0.66	6 GB	0.12 s	60 s
pdd 2.5D (w/o NL)	34.4±5.8			9 GB	0.54 s	29 s
pdd 3D (MIND)	44.7±4.6			22 GB	0.73 s	73 s
pdd 2.5D (MIND)	**44.8±4.9**	10.4	0.57	9 GB	0.54 s	29 s
NiftyReg	35.0					117 min

inference times sub-second on GPU, only 9 GB memory



PDD-Net results for Learn2Reg:
2nd best overall, Task3+4 unsupervised

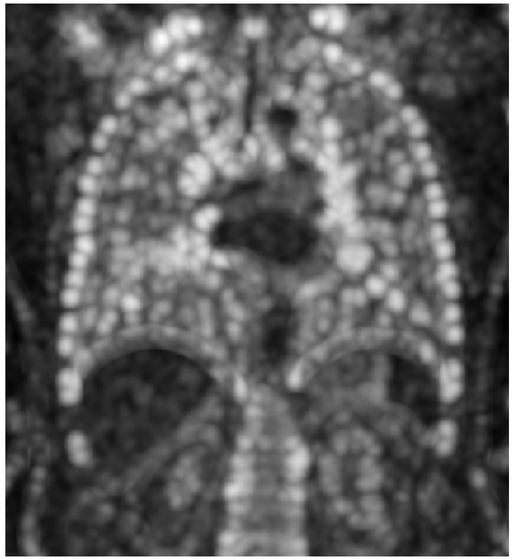
Ranks	Task01	Task02	Task03	Task04	Overall
Tony (LapIRN)	0.72	0.75	0.93	0.95	0.83
Lasse (PDD-Net)	0.94	0.85	0.69	0.74	0.80

for multimodal (Task1) + lung CT (Task2)
we used MIND-SSC (handcrafted features)

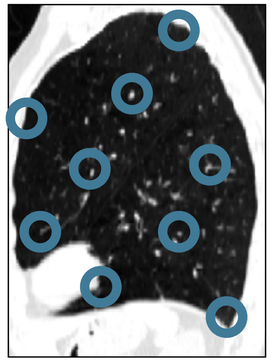
stapled 3D segmentation and mean intensity image after **registering all scans to one demonstrate qualitatively a high accuracy** (in all regions)

Concept for learning graph-based optimisation

Förstner Distinctiveness
→ **Keypoint extraction**

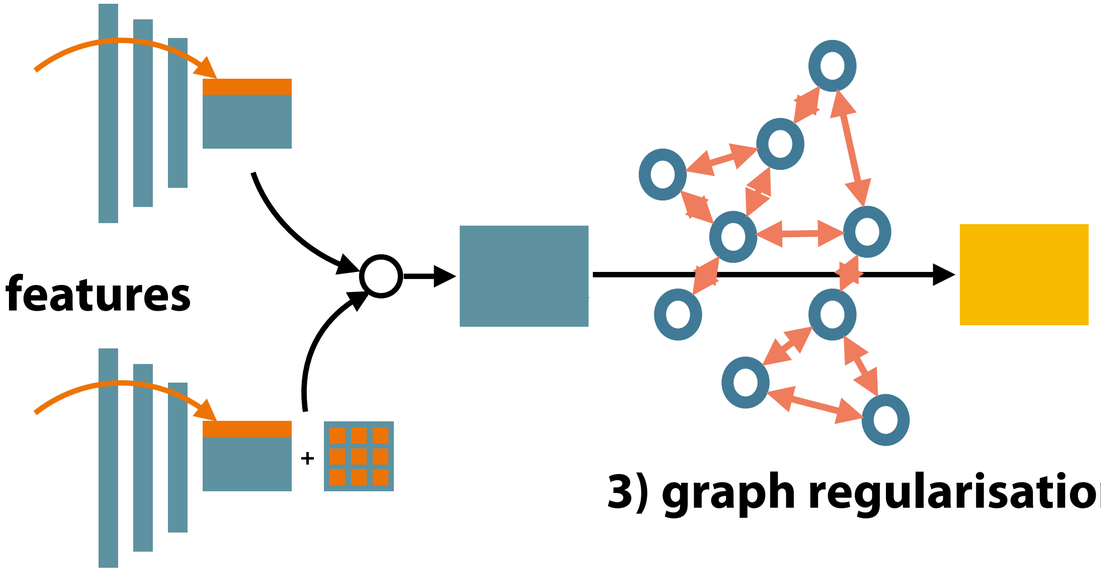


Inspiration Fixed I_F



Expiration Moving I_M

1) extract image **features independently for keypoints**
2) correlate them over **large search region**
3) **learn to regularise** motion jointly **on graph**



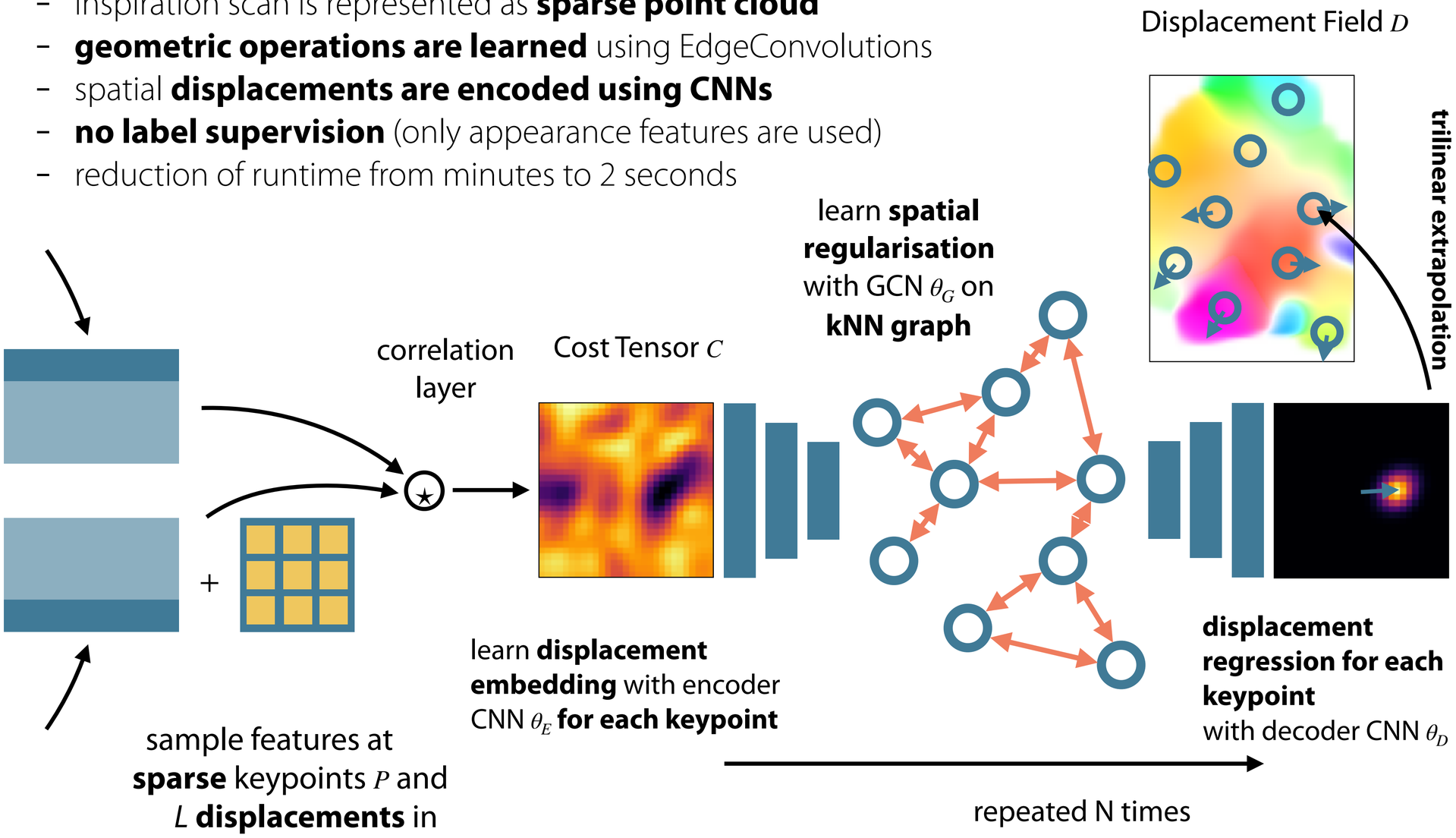
GraphRegNet has only ~33.000 trainable parameters

L Hansen and MP Heinrich: **GraphRegNet**: Deep Graph Regularization Networks on Sparse Keypoints for Dense 3D Medical Registration. **TMI in revision**

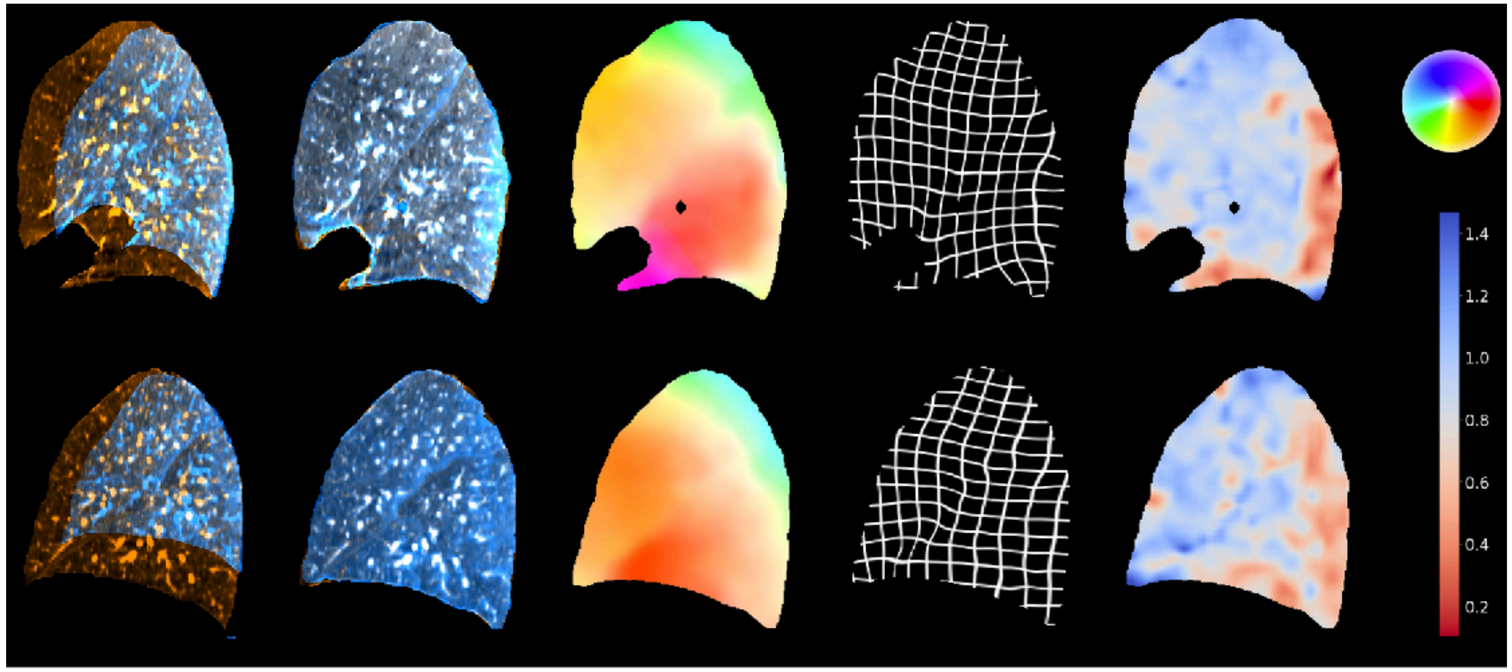
https://github.com/multimodallearning/graphregnet

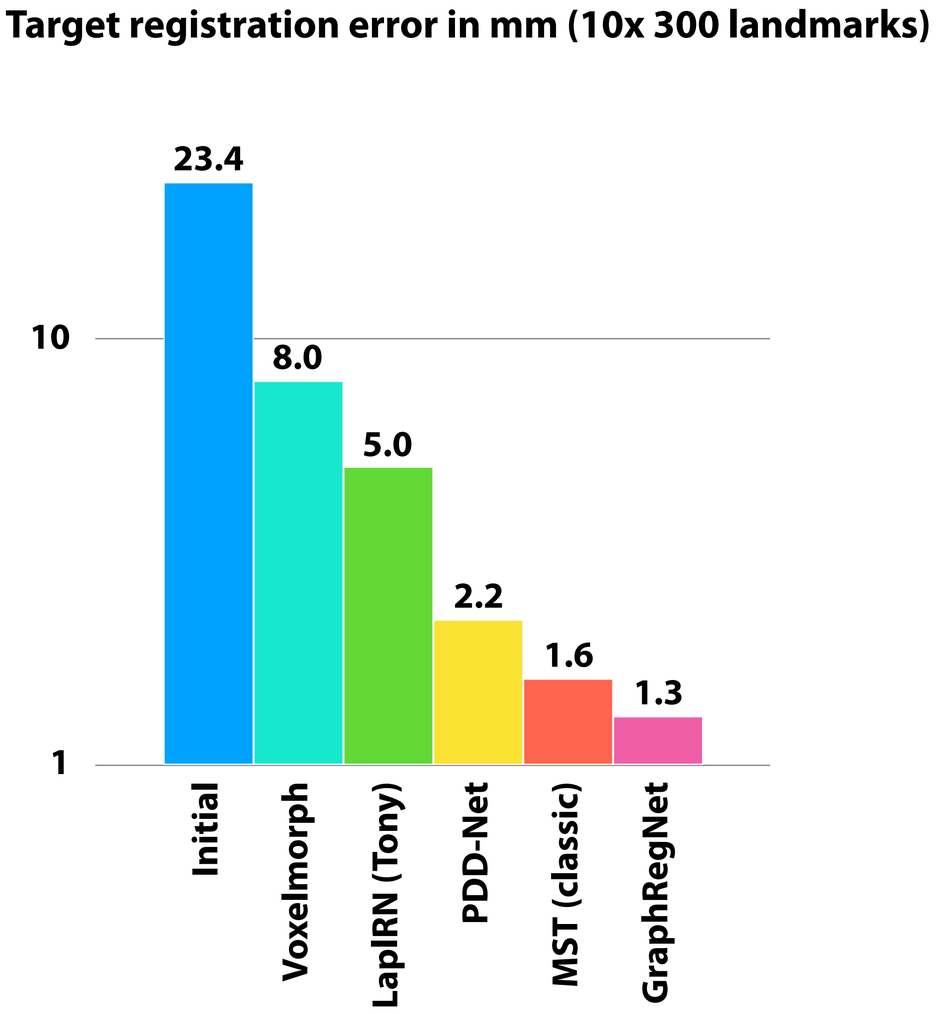
Geometric learning of graph-based optimisation

- inspiration scan is represented as **sparse point cloud**
- **geometric operations are learned** using EdgeConvolutions
- spatial **displacements are encoded using CNNs**
- **no label supervision** (only appearance features are used)
- reduction of runtime from minutes to 2 seconds



Results for DIR-LAB COPD dataset (inspiration/expiration)





- both our **discrete DL registration** frameworks PDD + GraphRegNet **outperform** U-Net/ResNet like architectures on lung registration
- learning **geometric regularisation** (EdgeConvs in GraphRegNet) **improve on** approximate **mean-field inference** (PDD-Net)
- feature learning seems of less importance (MIND-SSC sufficient)

Conclusions

Conclusions

- **impressive progress** over last few years **in medical DL registration**
- **established concepts** from DL segmentation or optical flow **are not directly transferable** new ideas: *multi-warping, two-stream, feature pyramids, synthetic data generation*, ..
- training DL-reg with **metric supervision requires** large dataset or **specifically designed loss computation:** *similarity pyramid, discrete nonlocal, advanced sampling*
- **fast registration possible w/o restriction** to single feed-forward → *instance optimisation*
- **probabilistic graph networks** enable geometric learning for **large deformations**
- unlike U-Net in segmentation: **few methods work out-of-the-box** for variety of registration tasks

Thank you for your attention - Q&A

Prof. Dr. Mattias Heinrich
- Medical Deep Learning Group
- Institute of Medical Informatics
- University of Lübeck, Germany

Thanks to my team in particular:
Lasse Hansen (PhD student geometric learning)
Hanna Siebert (PhD student DL registration)

Contact:
heinrich@imi.uni-luebeck.de
github.com/multimodallearning

see you in Lübeck for MIDL 2021

MIDL
Lübeck 2021
7 - 9 July 2021