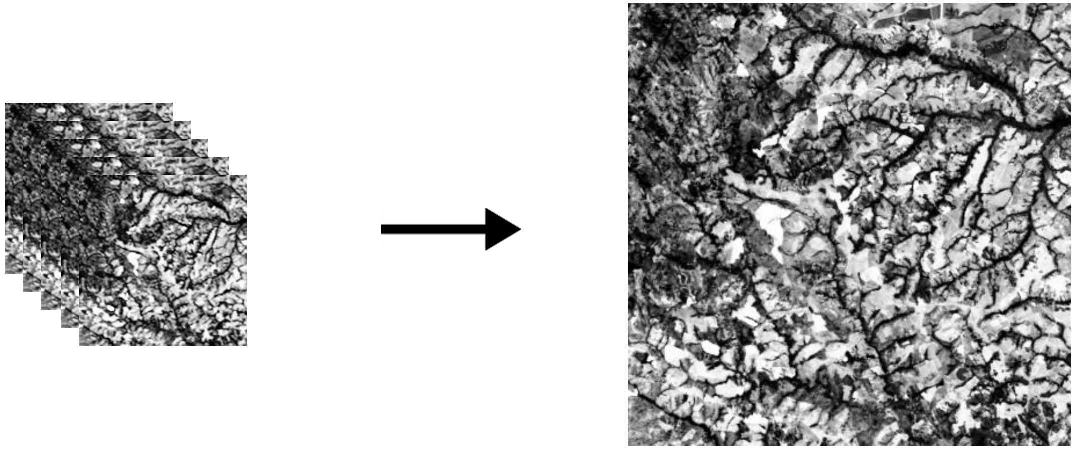
DeepSUM: Deep neural network for Super-resolution of Unregistered Multitemporal images

Andrea Bordone Molini, **Diego Valsesia**, Giulia Fracastoro, Enrico Magli
Politecnico di Torino

Kelvins day
11/9/2019

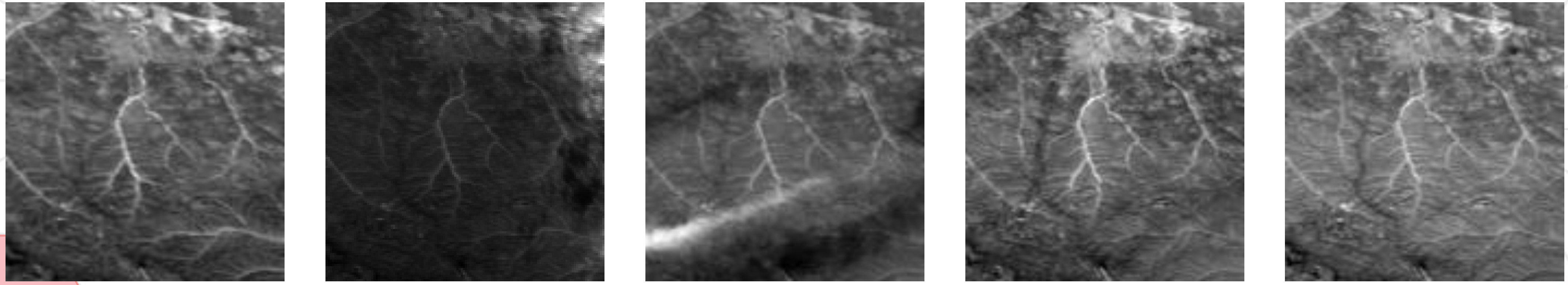
Deep learning for super-resolution

- **SISR**: single LR image
  - Exploit spatial correlation, prior knowledge about "natural images"
  - Widely studied (model-based, deep learning, …)
- **MISR**: multiple LR images to one HR image
  - Prior knowledge about "natural images" but also…
  - …correlation across multiple subpixel-shifted LR images
  - Very little work with deep learning, especially for remote sensing



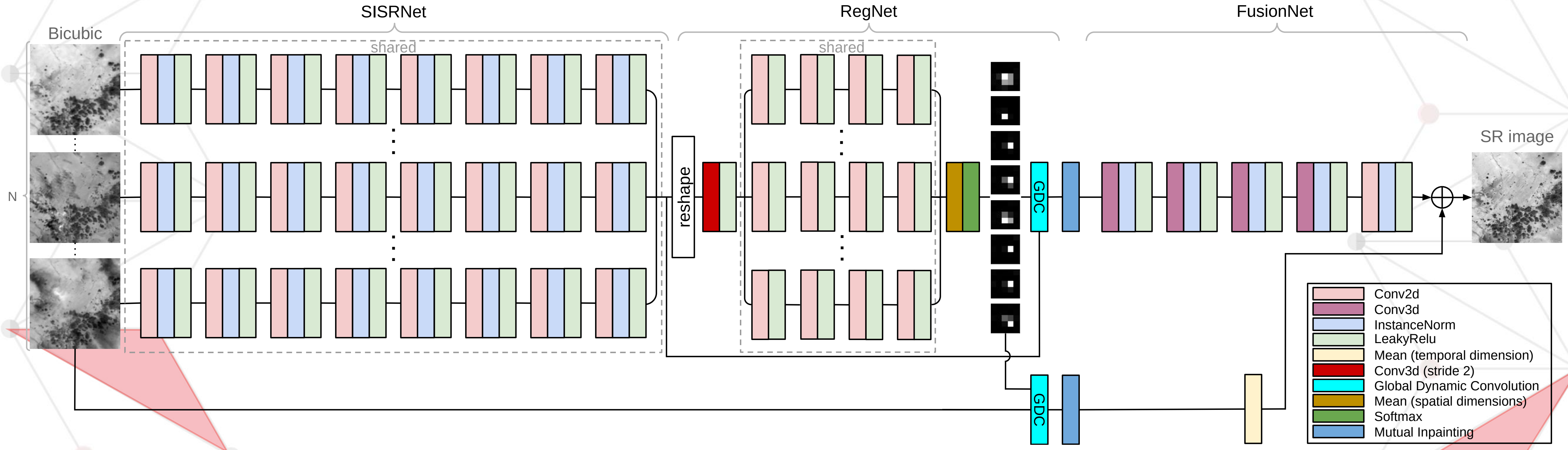
Dataset and issues

- Images from RED and NIR spectral bands at 300m and 100m resolution.
  - LR:**128x128 pixels**
  - HR: **384x384 pixels**
- Issues:
  - **Concealed regions**:
    - i.e. clouds, cloud shadows, ice, water, missing regions
  - **Change in brightness and content:**
    - the images have been recorded within a time window of 30 days
  - **Unregistered images**
- Masks: regions with unreliable pixels.



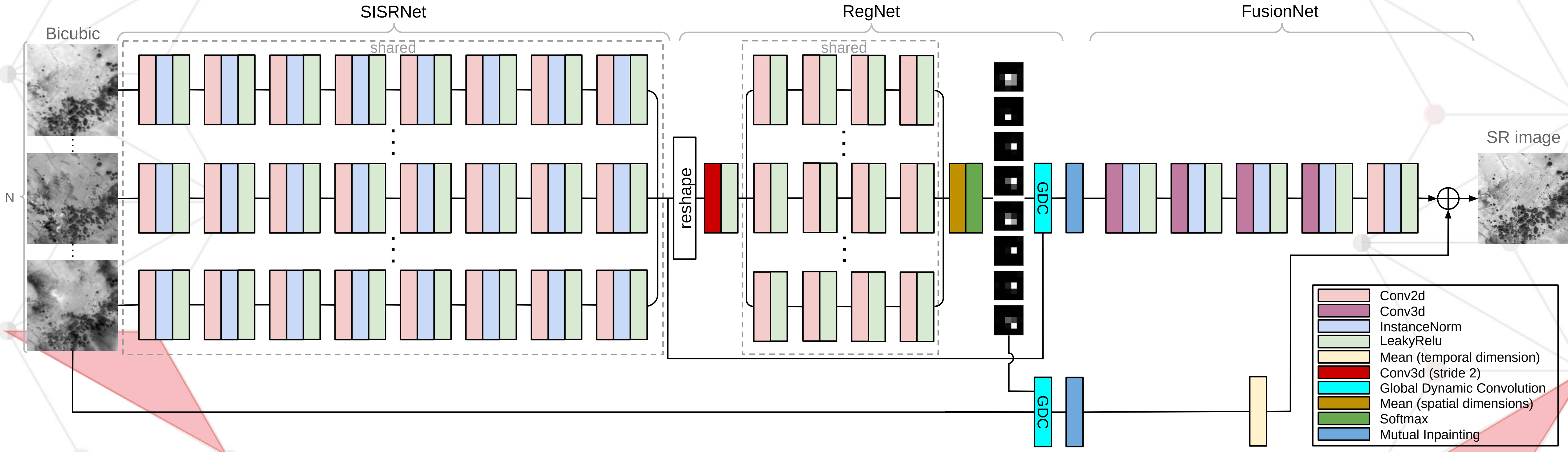
DeepSUM

- **DeepSUM** relies on a single residual CNN with three main stages:
  - **SISRNet**: shared 2D convolutions to extract high-dimensional features;
  - **RegNet**: subnetwork proposing registration filters from the high-dimensional features;
  - **FusionNet**: 3D convolutions for slow fusion of the features from multiple images.
- **Tricks**: residual, inpainting of masked regions



DeepSUM

- Pretrain SISRNet and RegNet separately with a task-specific objective
- Finetune everything with the super-resolution objective



Loss

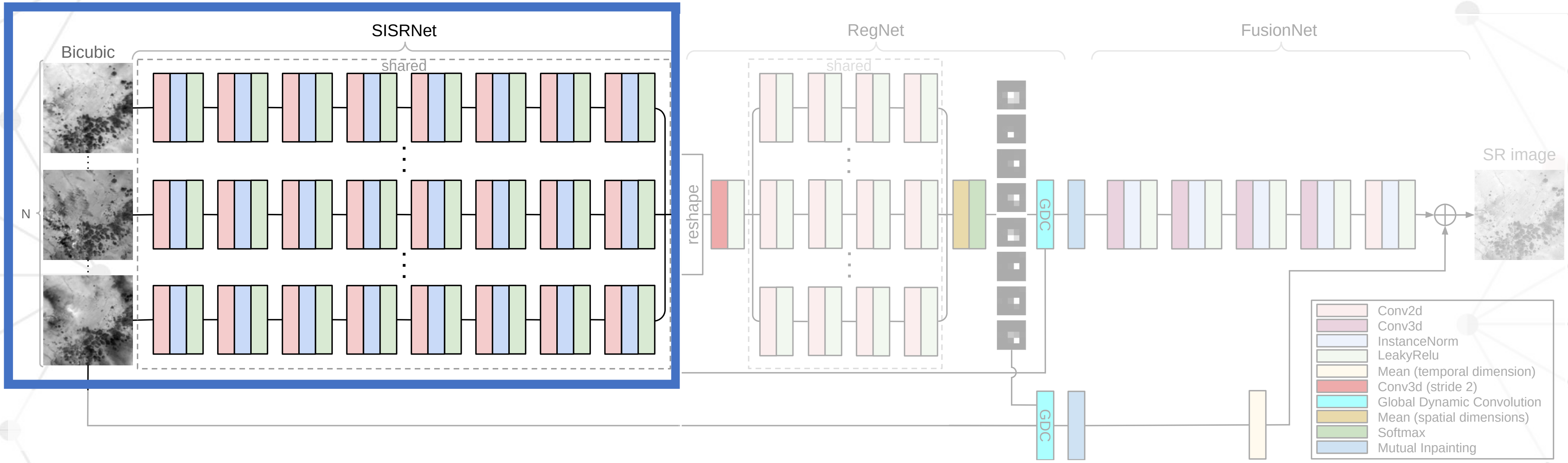
- **Robustness** to many factors:
  - The brightness of the HR image may be different from that of the SR image
  - The SR image and the HR image may be unregistered.

- Compensate absolute brightness difference
- Test all SR-HR alignments and select best
- Same as challenge scoring function

$$L = \min_{u,v \in [0,2d]} \| I^{\mathrm{HR}}_{u,v} - (I^{\mathrm{SR}}_{\mathrm{crop}} + b) \|^2$$

$$b = \frac{1}{(rW-d)(rH-d)} \sum_{x,y} (I^{\mathrm{HR}}_{u,v} - I^{\mathrm{SR}}_{\mathrm{crop}})$$

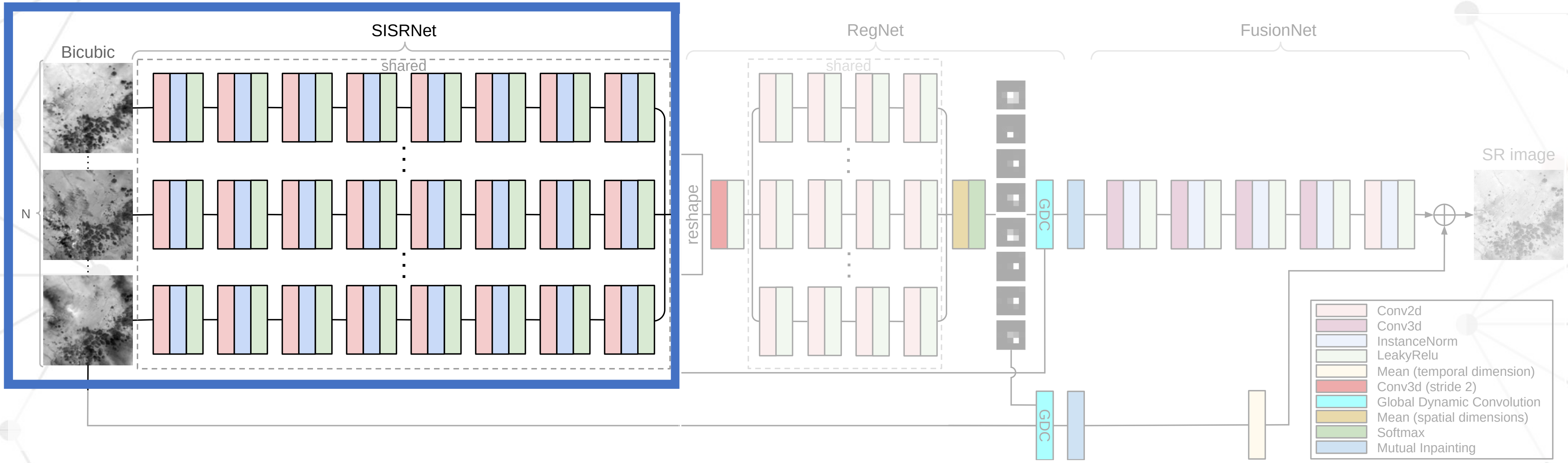
SISRNet



Goals:

- create a high-dimensional feature space suitable for downstream tasks
- exploit intra-image correlation to perform single-image SR

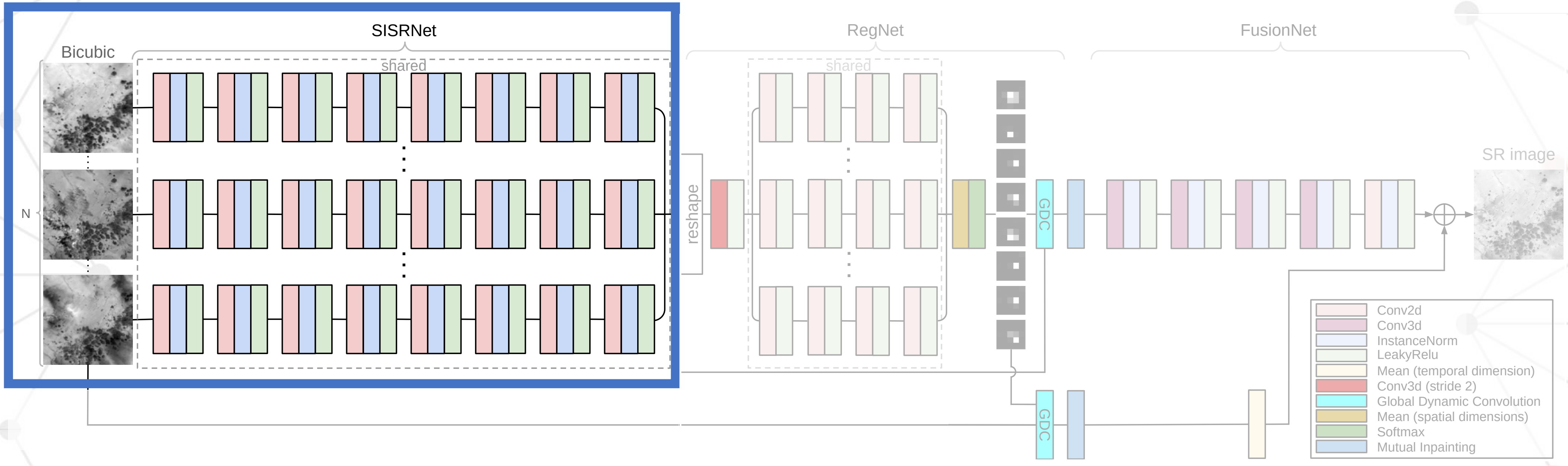
SISRNet



Architecture:

- Input: bicubically upsampled images
- Conv2D – InstanceNorm – LeakyRelu shared across images
- can be improved integrating latest works on SISR
- InstanceNorm > BatchNorm : contrast and brightness variations among training images within a batch may cause instability

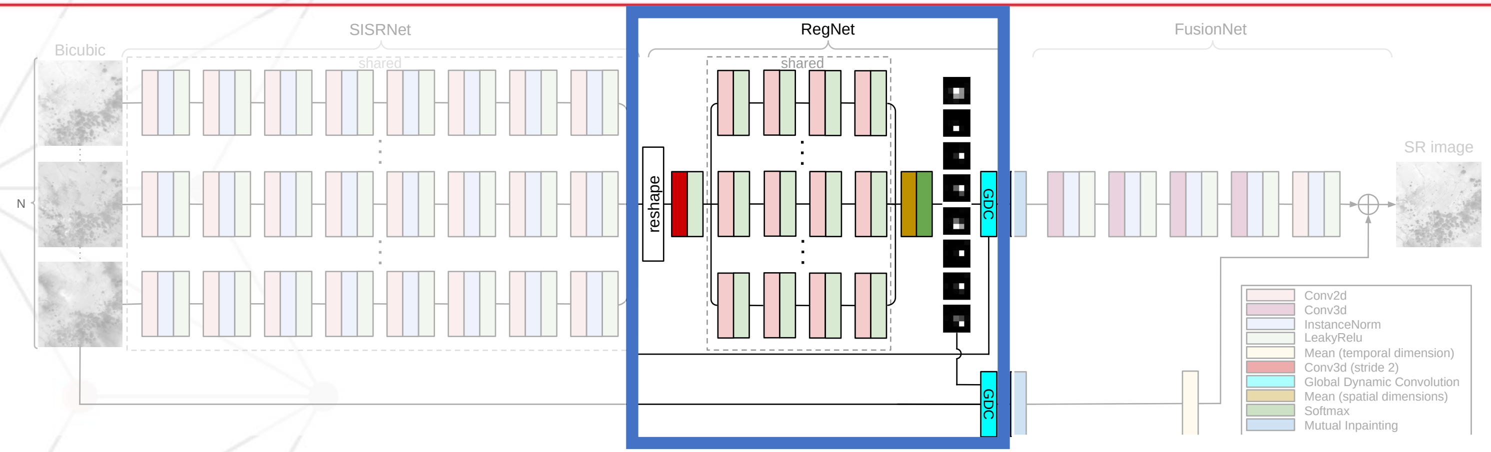
SISRNet

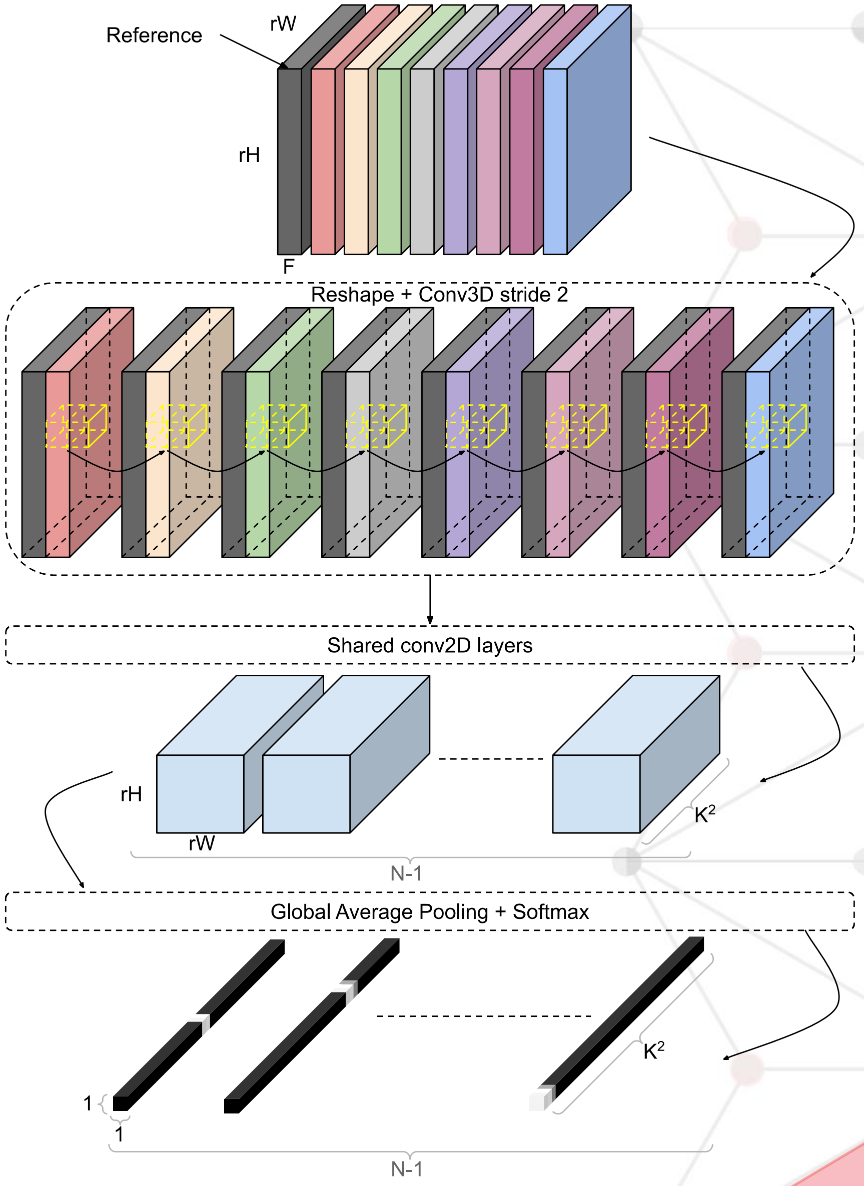


Pretraining:

- add a final convolutional layer to map from F features to 1 channel
- train with SR loss function
- remove final layer: we want to keep working in the feature space in the full architecture

RegNet





- **Register** the *N* image representations coming from SISRNet
- Registration relative to the first image
- Two sub-blocks:
  - CNN: outputs *N-1* registration filters (9x9)
  - global dynamic convolutional layer (GDC): applies the filters to the feature maps

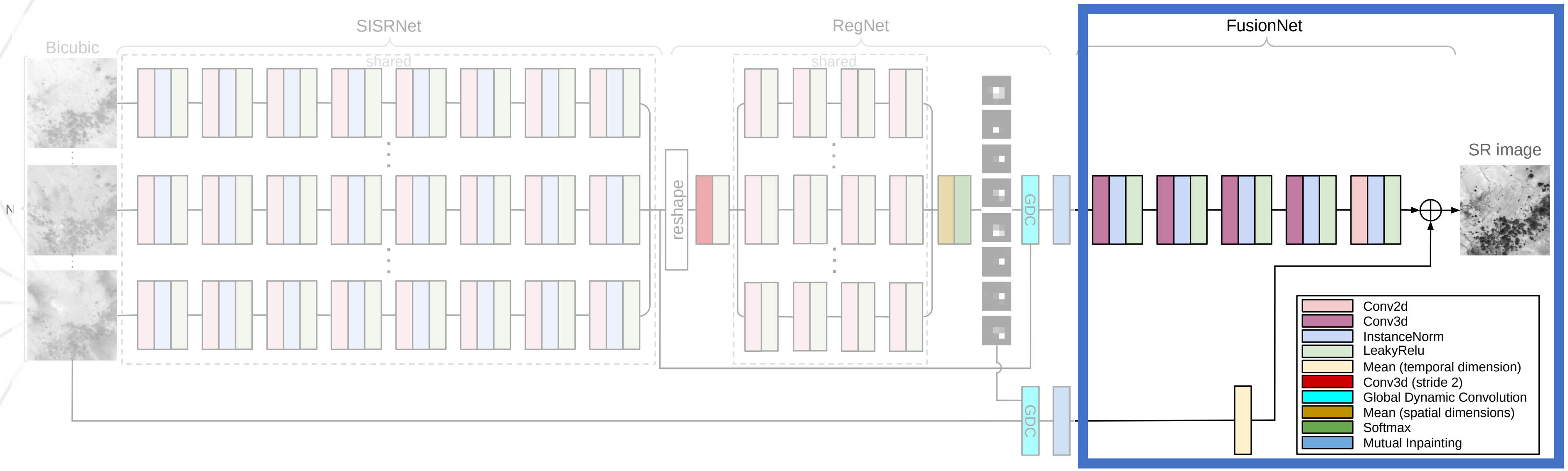
RegNet - pretraining

- Generate a dataset of high-dimensional features
  - Pass the training images through SISRNet
- Generate N-1 versions of the first image with synthetic shifts
- Train with cross-entropy objective between output filters and ideal filters
  - “Ideal” registration filters are delta functions centered at the right shift

RegNet – Pros and Cons

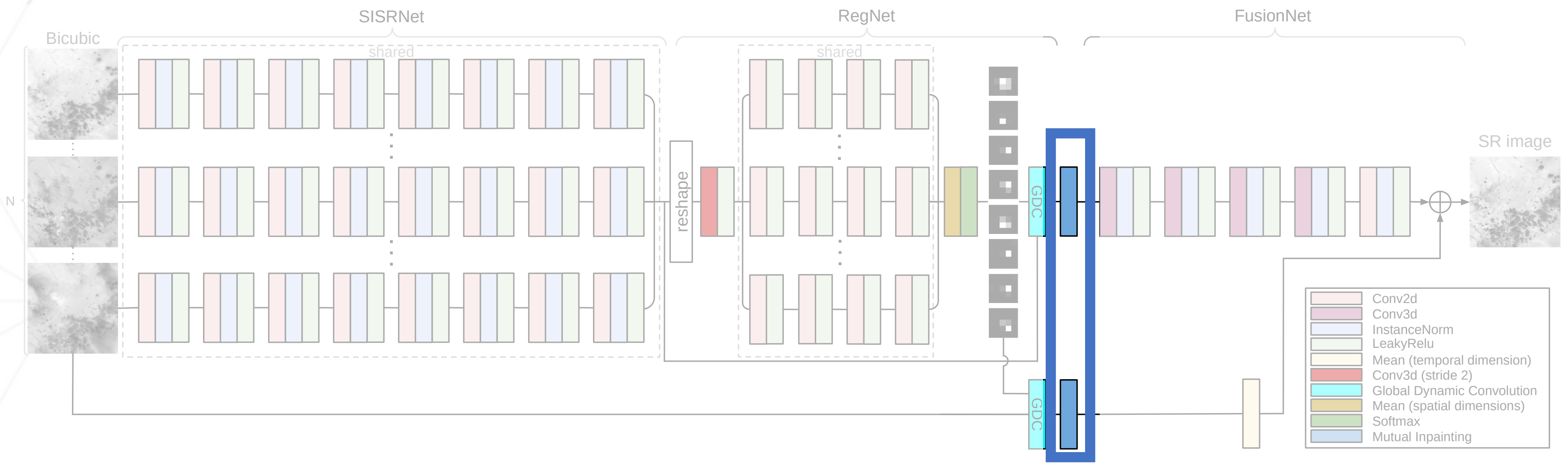
Image cross-correlation	Fixed kernel – learned shifts	RegNet
Shift = max of cross correlation in image space Fixed interpolation kernel	Shift = predicted by a neural net in feature space Fixed interpolation kernel	Shift and interpolation kernel predicted by a neural net
Very sensitive to image content	Robust to image content	Robust to image content
Non-adaptive interpolator (artifacts)	Non-adaptive interpolator (artifacts)	Adaptive interpolator (reduces artifacts)
Does not require training	Easy to train (2 output parameters)	Complex to train for large shifts

FusionNet



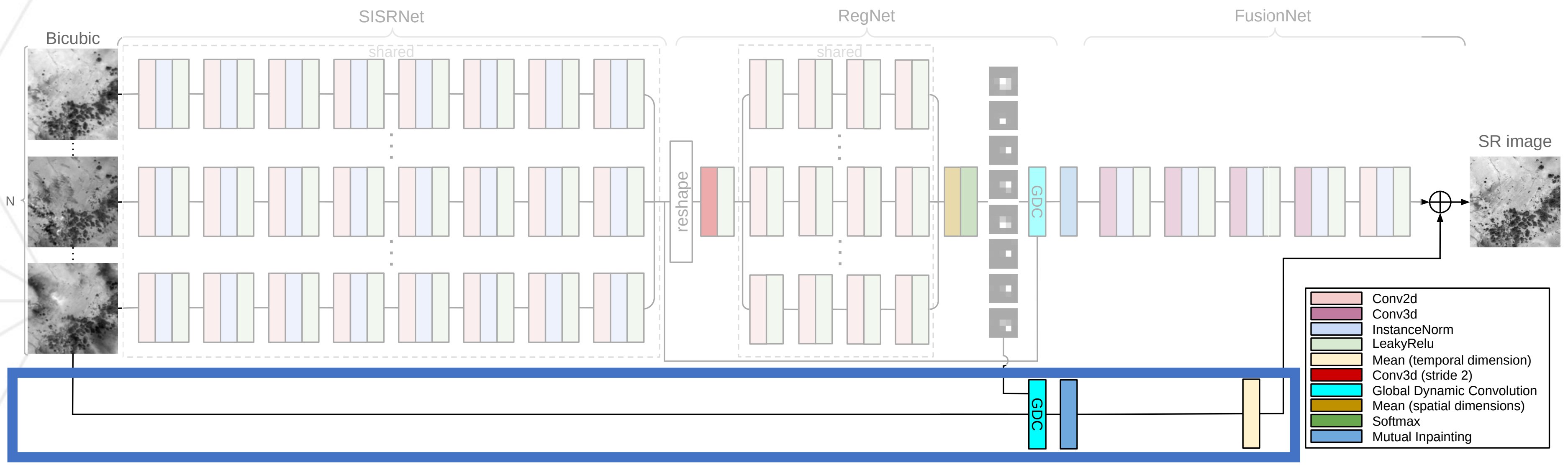
- FusionNet subnetwork progressively **fuse the N registered outputs** delivered by RegNet in a “slow” manner in the feature space
  - 3D convolutional layers without padding

Mutual inpainting



- Images contain masked areas with unreliable values (clouds, corruptions, …)
- Reliable fusion must discard such areas
- Fill unrealiable areas of feature maps with feature values from other images
- (Masks are also aligned with registered feature maps by GDC)

Residual



- Global input-output residual connection: learn how to fix bicubic upsampling
- Bicubically-upsampled input
  > Registration with the dynamic filters
  > Inpainting
  > Average

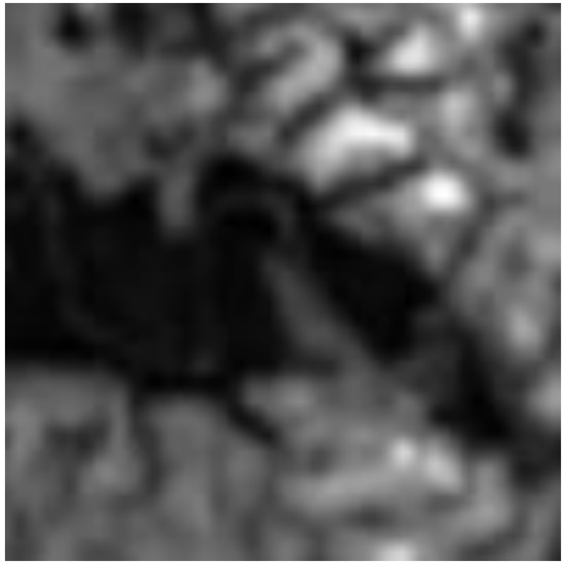
Testing with more images

- DeepSUM requires a fixed number of input images: N
- More than N images can be used by combining multiple N-images estimates
- In the challenge we used a sliding window approach
  - Sort test images by increasing amount of masked pixels
  - SR estimate from a window of N=9 images
  - Move window by 1 image

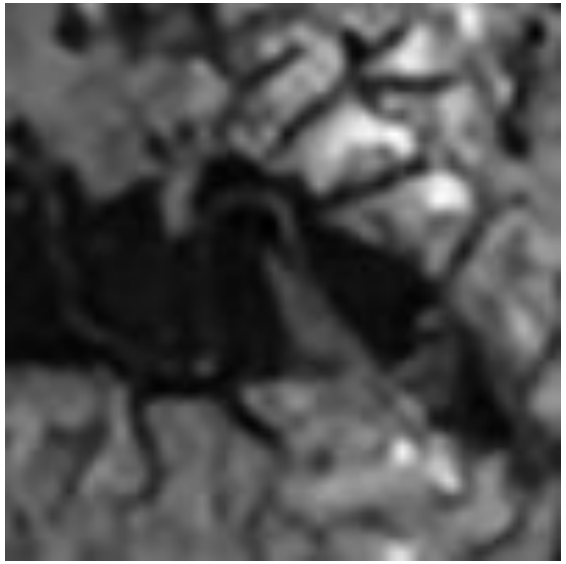
Qualitative Results (RED)

LR

Bicubic+Mean : 46.32 dB

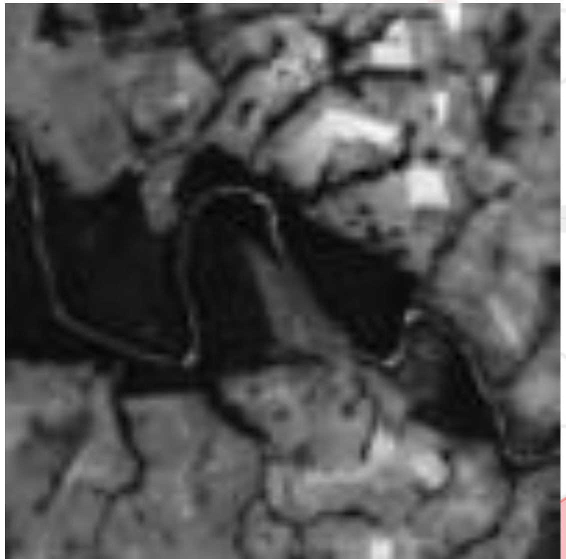


SISR+Mean: 48.33 dB



DeepSUM: 49.89 dB

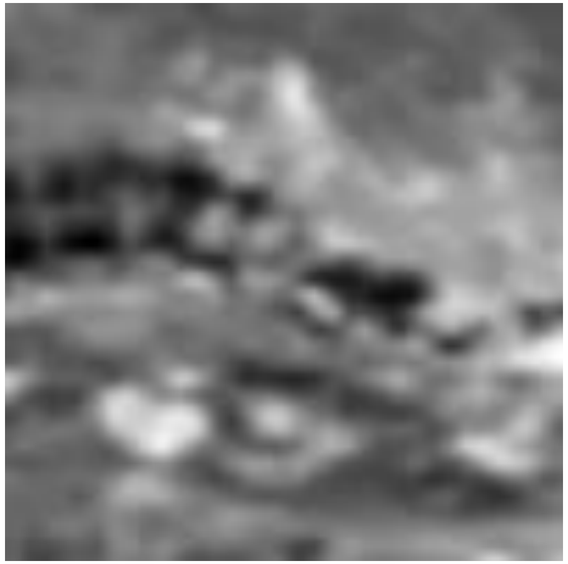
HR



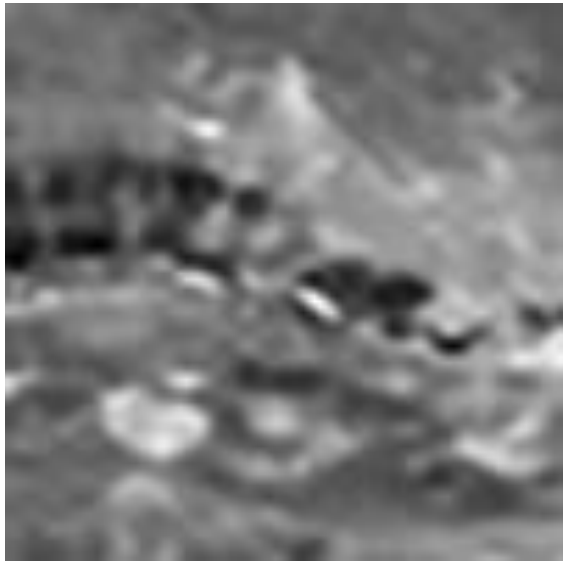
Qualitative Results (NIR)

LR

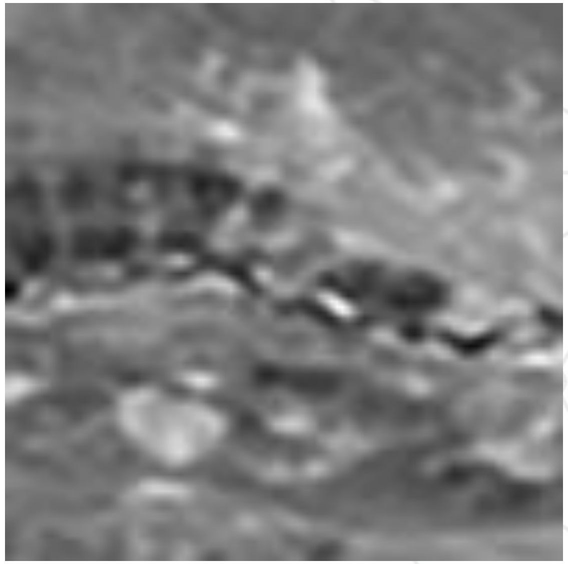
Bicubic+Mean : 47.71 dB



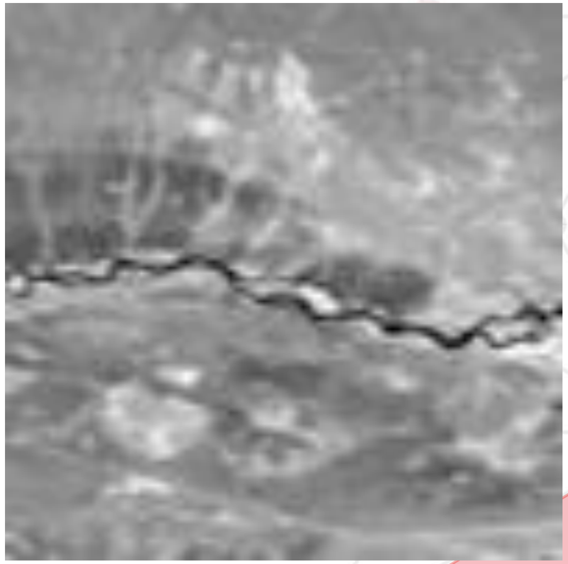
SISR+Mean: 48.68 dB



DeepSUM: 50.82 dB



HR



Quantitative Results

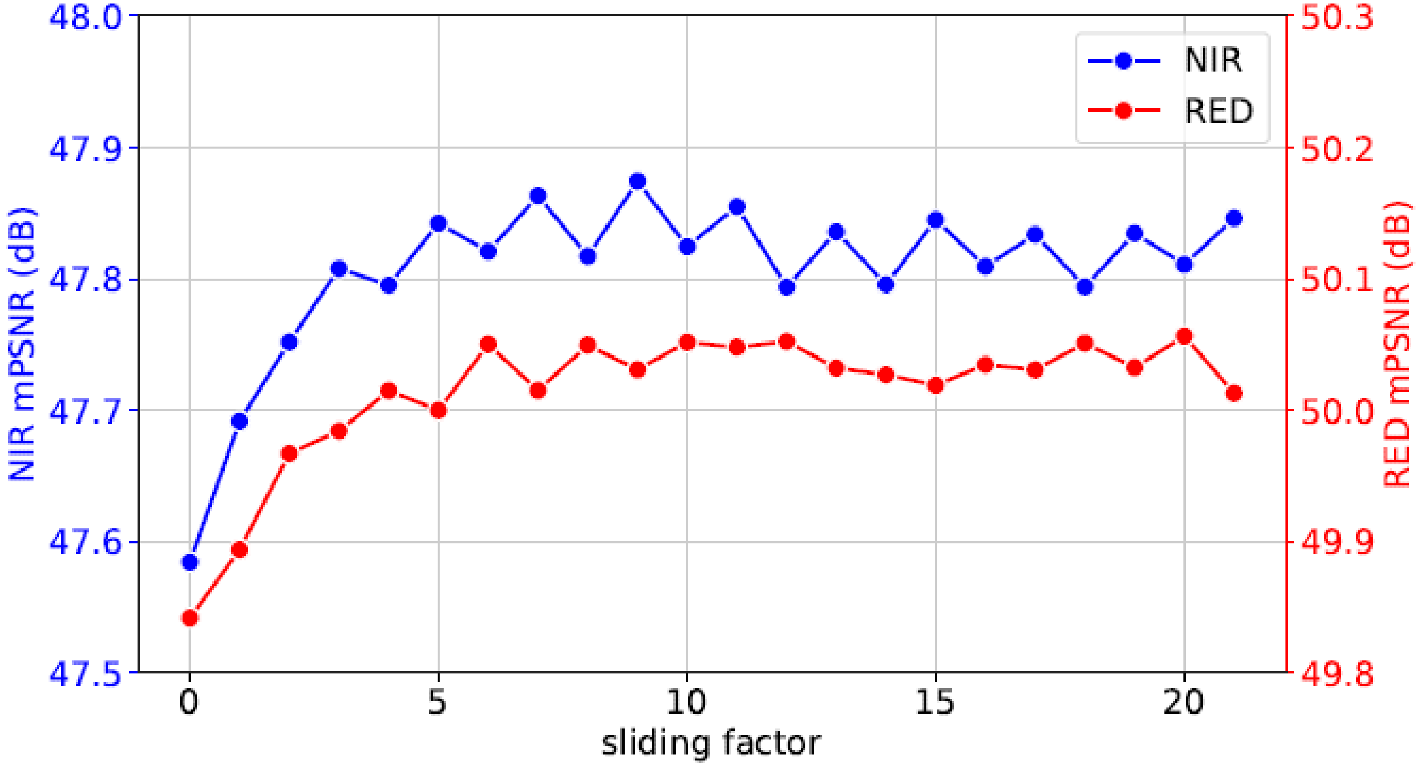
TABLE I
AVERAGE MPSNR (DB) - REGNET PERFORMANCE

	Proposed without RegNet	**Proposed with RegNet**
NIR	47.68	**47.84**
RED	49.87	**50.00**

TABLE II
AVERAGE MPSNR (DB)

	Bicubic+Mean	IBP [36]	SISR	SISR+Mean	DUF [58]	**DeepSUM**
NIR	45.69	45.96	45.56	46.57	47.06	**47.84**
RED	47.91	48.21	48.20	48.82	49.36	**50.00**

Quantitative Results



- Using more than 9 LR images helped
- Using more than the 15 least masked images did not improve performance

Future work

- **SISRNet**
  - **Non-local models**: recent work on extending convolutional layers to include non-locality. See:

    D. Liu, B. Wen, Y. Fan, C. C. Loy, T. S. Huang, «***Non-Local Recurrent Network for Image Restoration***», NeurIPS 2018

    D. Valsesia, G. Fracastoro, E. Magli, «***Deep Graph-Convolutional Image Denoising***», arXiv:1907.08448

Journal reference

A. Bordone Molini, D. Valsesia, G. Fracastoro, E. Magli –”***DeepSUM: Deep neural network for Super-resolution of Unregistered Multitemporal images***”
under review IEEE Transactions on Geoscience and Remote Sensing

arXiv:1907.06490 preprint

Code: github.com/diegovalsesia

Poster at IEEE WHISPERS 2019

www.ipl.polito.it

Follow us on Twitter: @IPLPolito