Semantic Segmentation for Geophysical Data

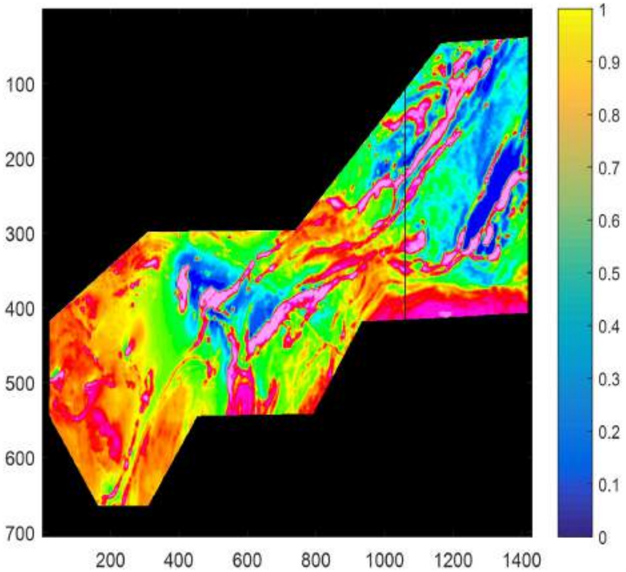
Xiaoijn Tan, Eldad Haber

Please check the link to see the paper related to this work sample.

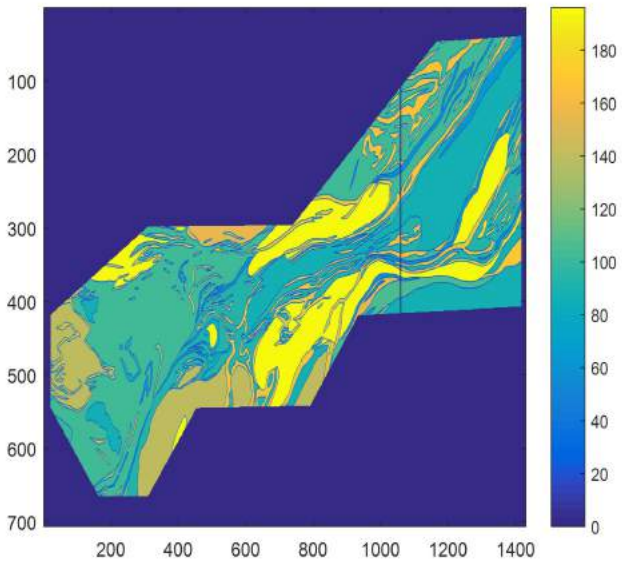
Motivation

Question addressed:

How to extract relevant geological information from geophysical data, leading to a geological unit map ?



Segmentation for Magnetic Data

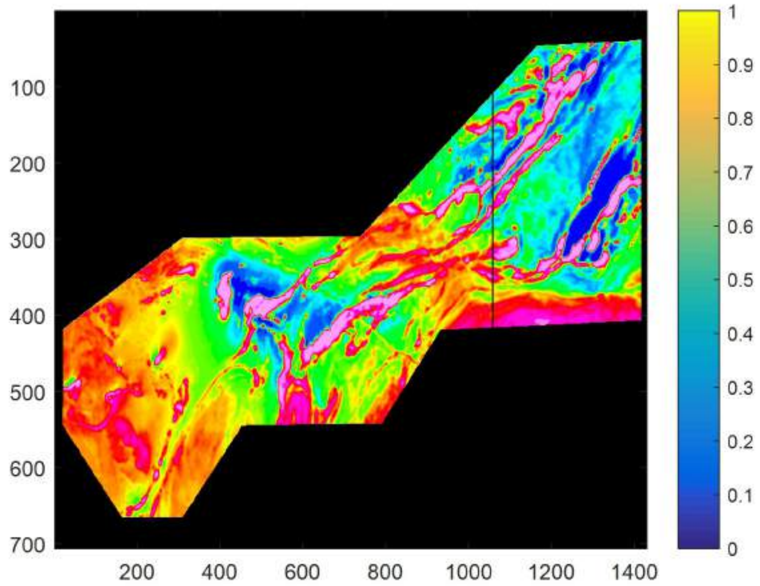


different magnetism — Different rocks

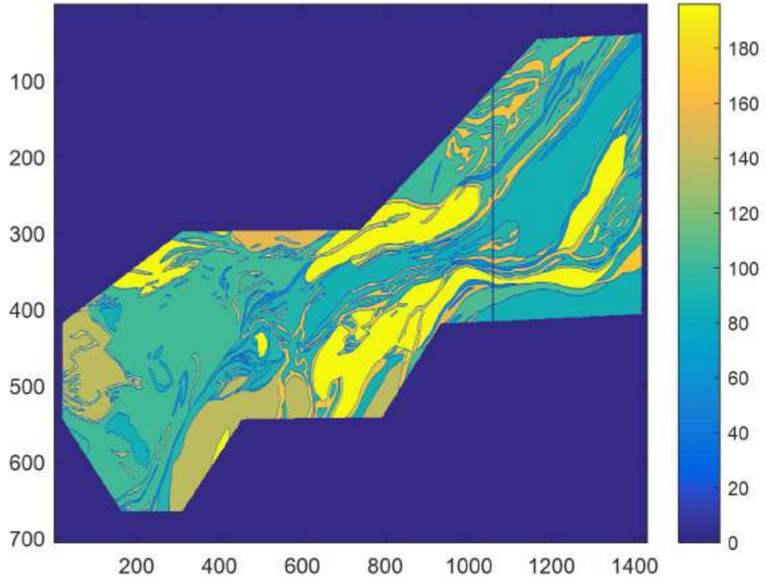
changes in the magnetic fields — changes in geological units

Motivation

- **To date**, segmentation for geophysical data is typically done **manually**
- **Limitations**
    - Experienced geologists needed
    - Time and effort consumed



Done by Geologists



Motivation

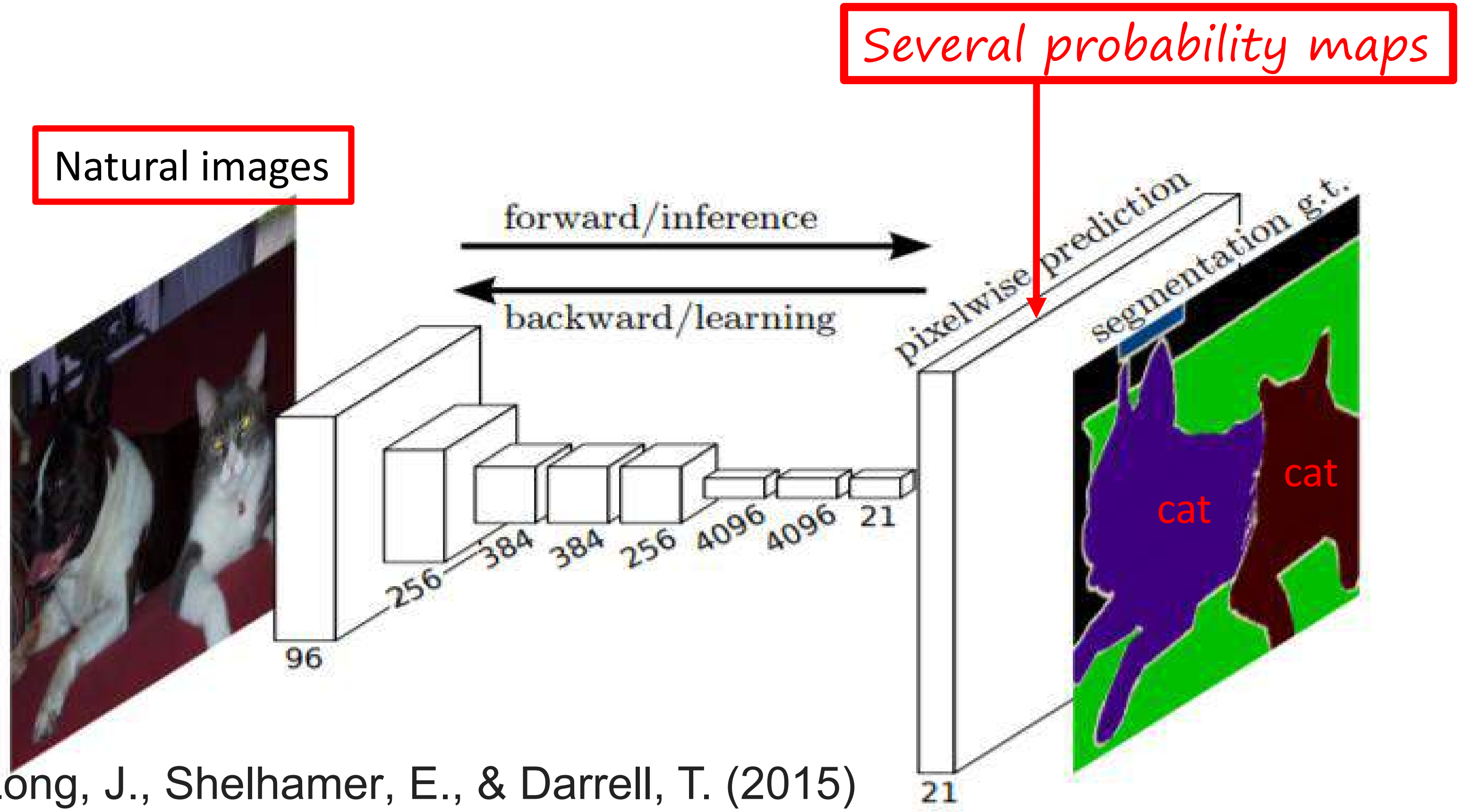
Question addressed:

Can we borrow ideas from other fields to overcome these limitations?

Motivation

Question addressed:

Can we borrow ideas from other fields to overcome these limitations? Yes! Use CNNs in semantic segmentation

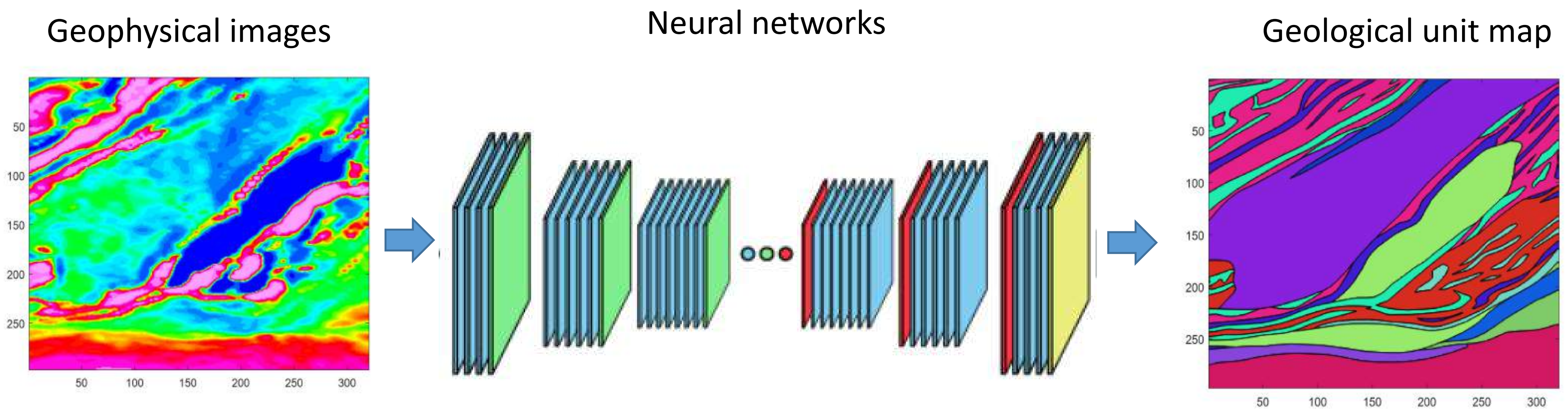


- **Advantages**
  - automatically
  - no experience required
  - Fast, without effort.

Long, J., Shelhamer, E., & Darrell, T. (2015)

Basic ideas

- attempting to use similar technology (CNNs) for geological segmentation



Challenges

- Geophysical data vs. natural images

  - Data inconsistency (not discussed here since it is not related to machine learning problems)

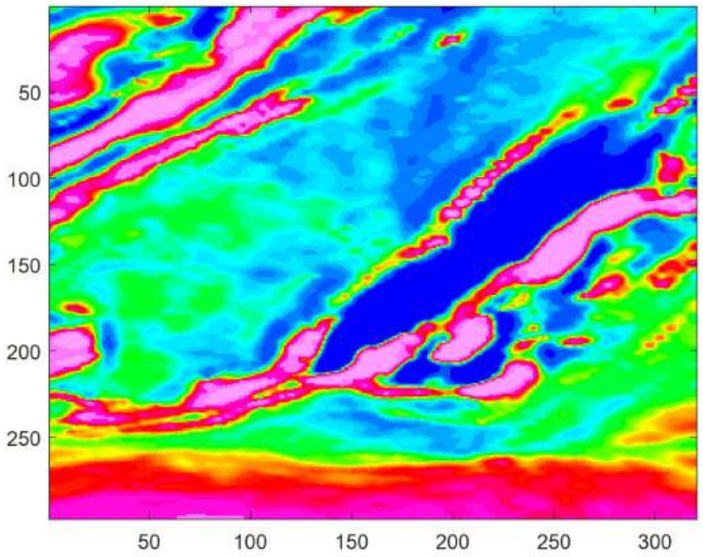
  - Labeled Data scarcity

  - Data complexity

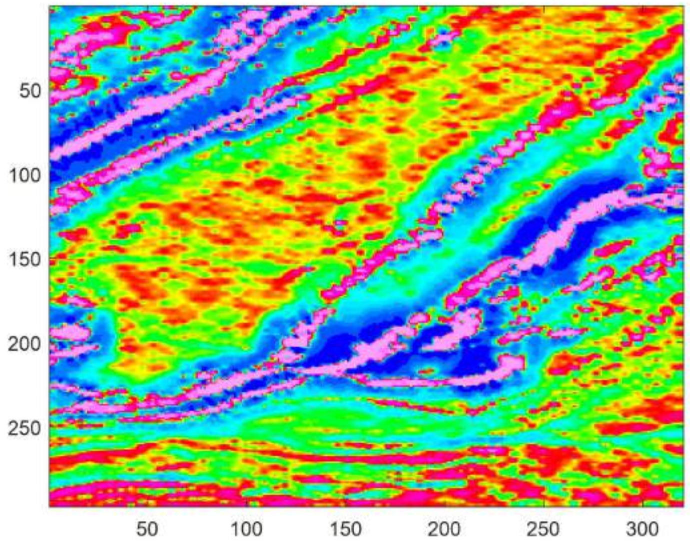
Challenges

- **Labeled Data scarcity**
  - E.g. magnetic data

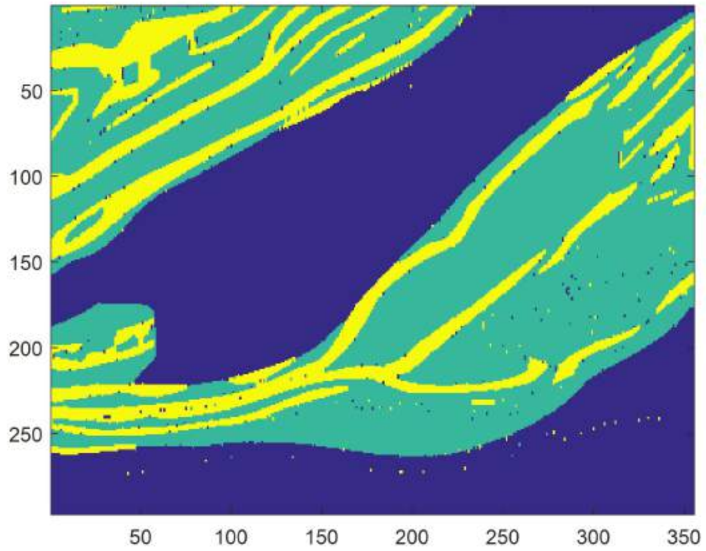
labels



- Sufficient number
- From geophysical surveys



- Sufficient number
- From computations



- Insufficient number
- *require efforts by geologists*

Challenges

- Labeled Data scarcity
  - E.g. magnetic data
- Insufficient number of pixel-wise labels for training
  - overfitting and poor generalization

Challenges

- Labeled Data scarcity
  - E.g. magnetic data
- Insufficient number of pixel-wise labels for training
  - overfitting and poor generalization
- Data augmentation
  - Synthetically creating new samples from original training dataset

Challenges

- Geophysical data vs. natural images

  - Labeled Data scarcity → Data augmentation

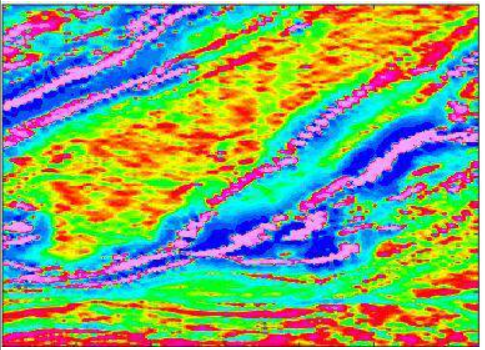
  - Data complexity

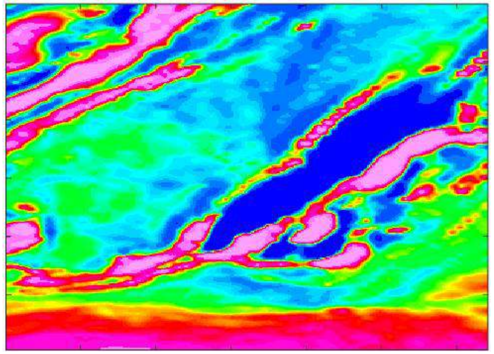
Challenges

- Geophysical data vs. natural images
  - ➢ Labeled Data scarcity → Data augmentation
  - ➢ Data complexity

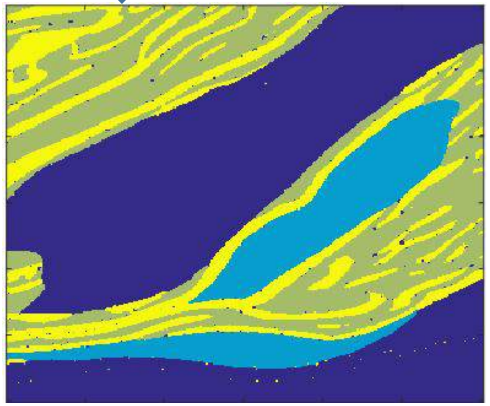
Challenges

- Data complexity
  - more complex image-label relationship

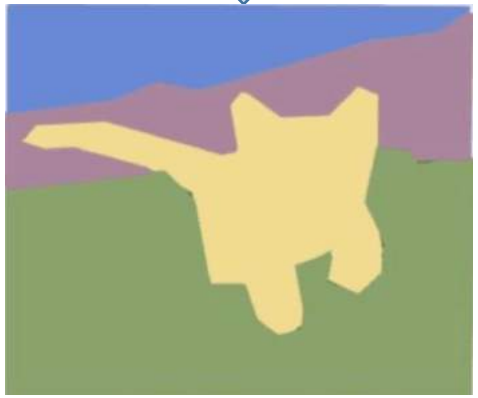




- Experienced geologists



- Any schoolchild quite likely to know how to segment



Challenges

- Geophysical data vs. natural images

  - Labeled Data scarcity → Data augmentation
  - Data complexity → New Neural Network architecture

Methods

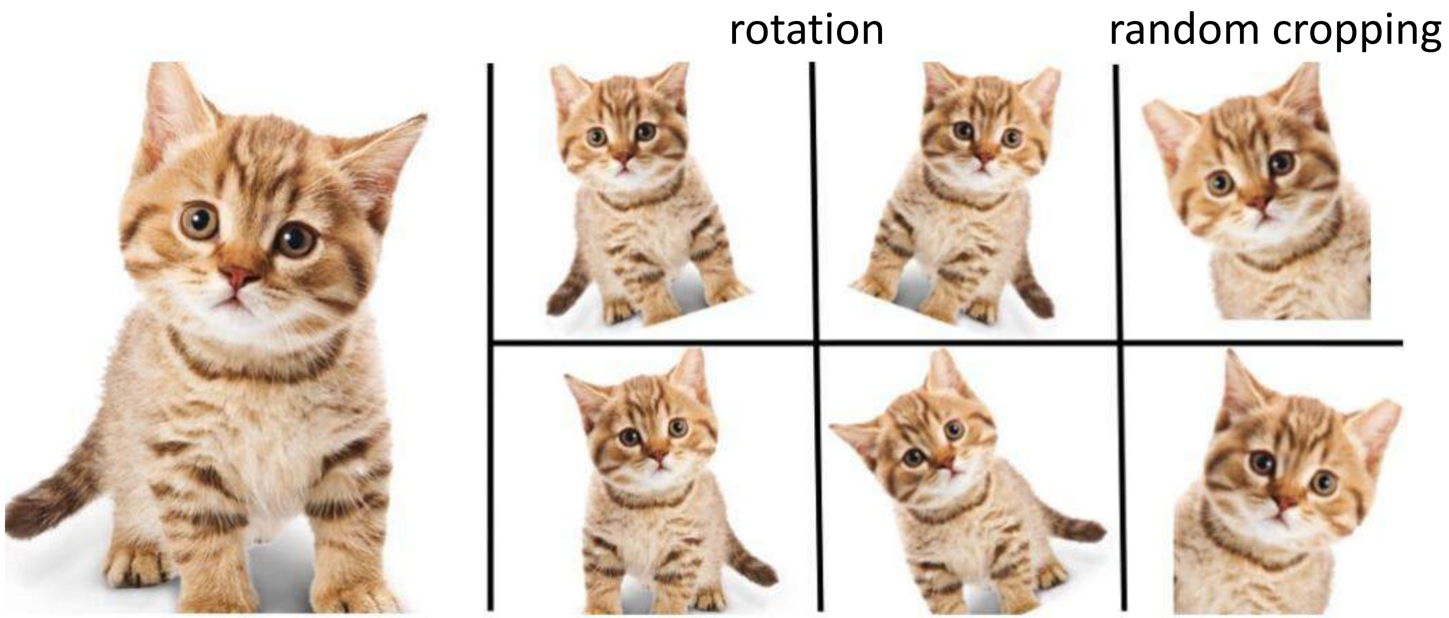
- Data augmentation

- Supervised learning

    - Next will show how to implement these methods in magnetic data

Traditional data augmentation

- Traditional data augmentation
  - mirroring, random cropping, rotation, shearing and color shifting
  - fail to take the special characteristics of geophysical data into account
    - ➢ may be geologically infeasible
    - ➢ may ruin spatial continuity
    - ➢ Simply copy the original data

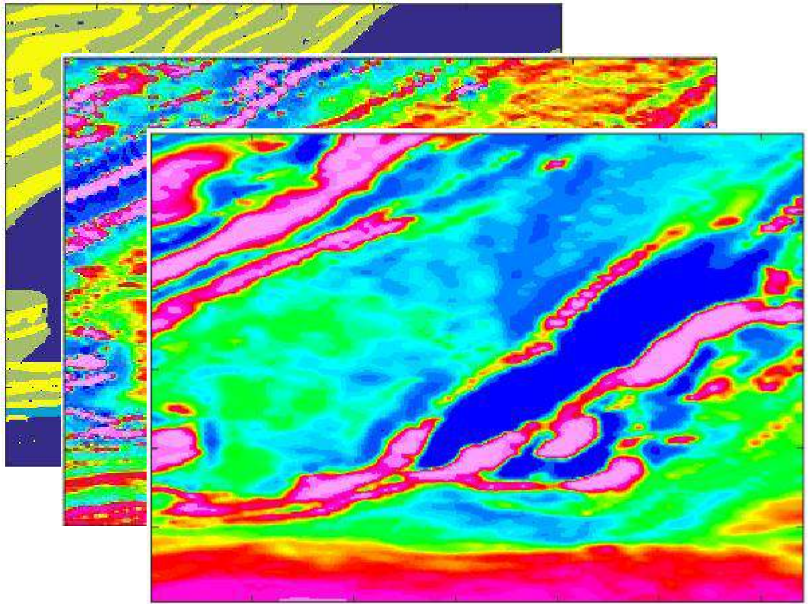


Data augmentation

Question addressed:

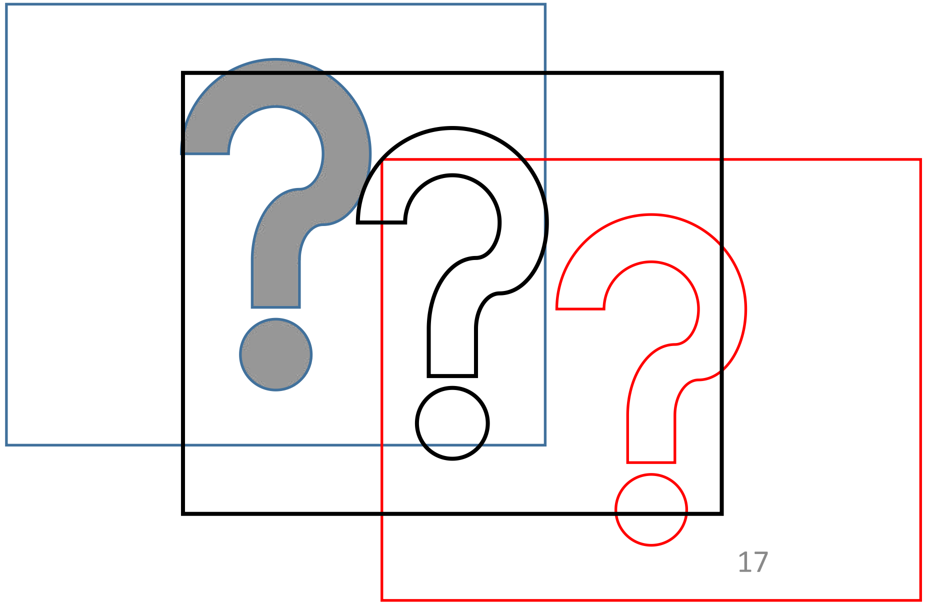
How to synthesizing geologically feasible samples that share the same spatial continuity with the original data?

Training images



Data augmentation

- ❑ Geologically feasible
- ❑ Same spatial continuity

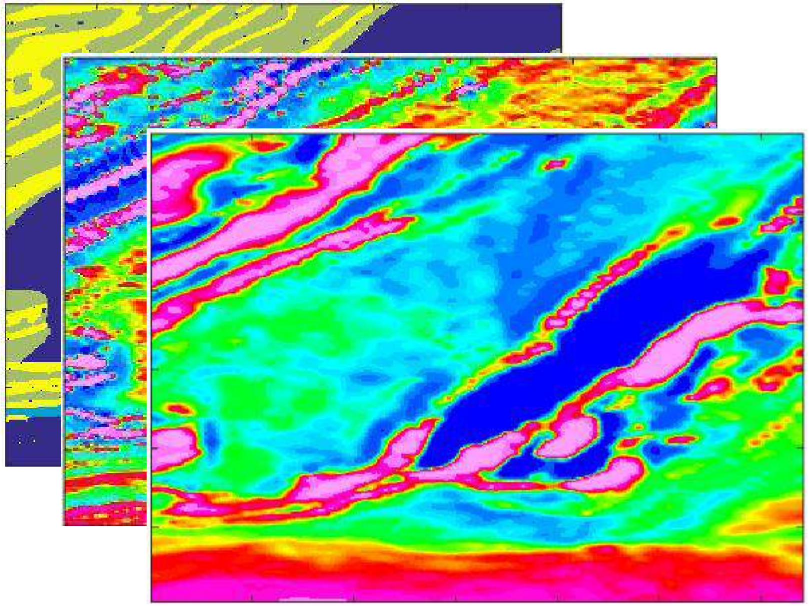


Data augmentation

Question addressed:

How to synthesizing geologically feasible samples that share the same spatial continuity with the original data?

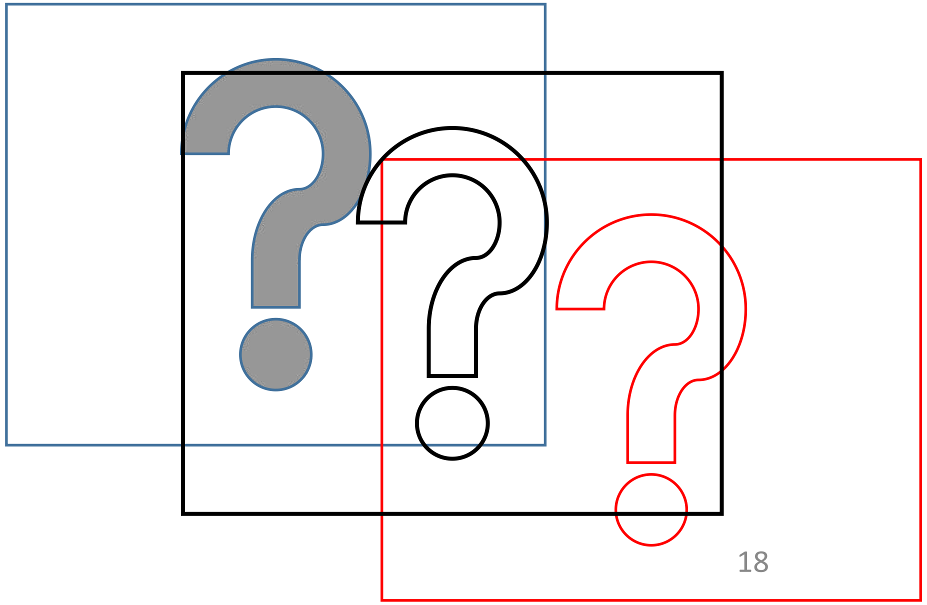
Training images



Data augmentation

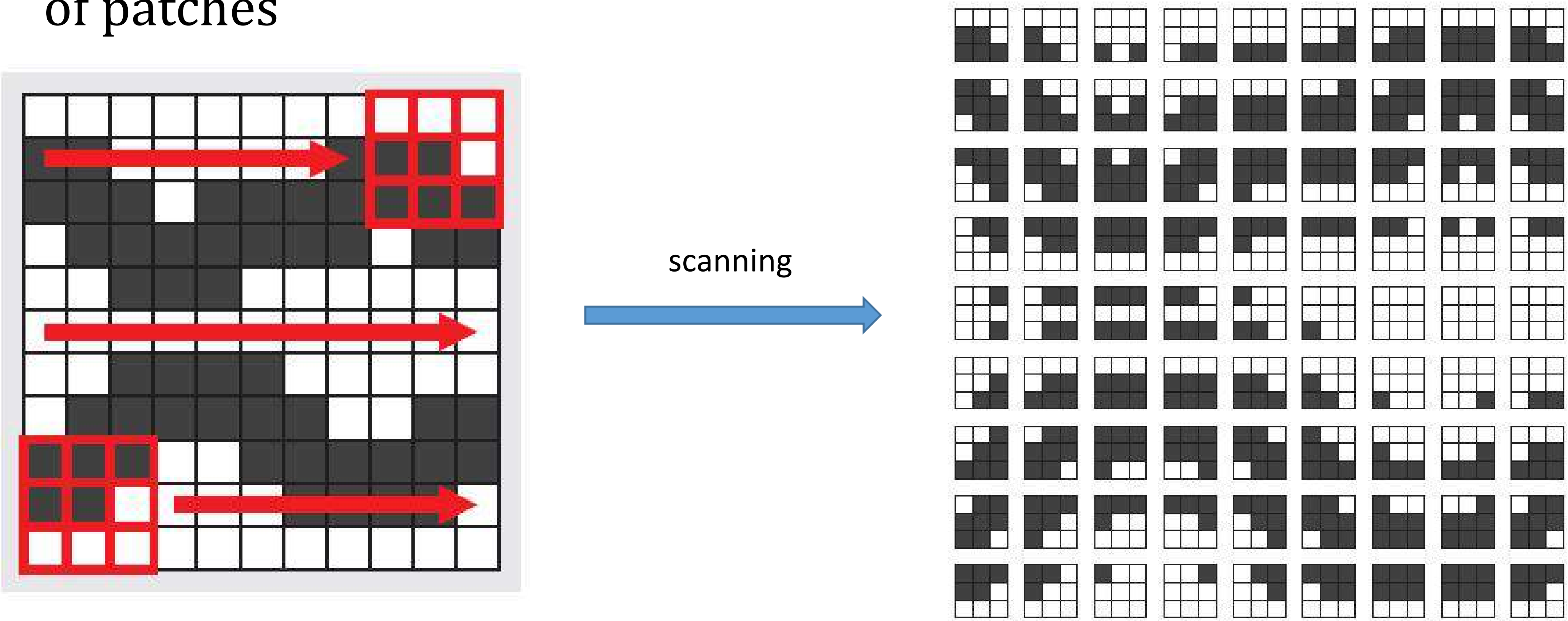
Multi-point geostatistics

- ❑ Geologically feasible
- ❑ Same spatial continuity



Basic idea

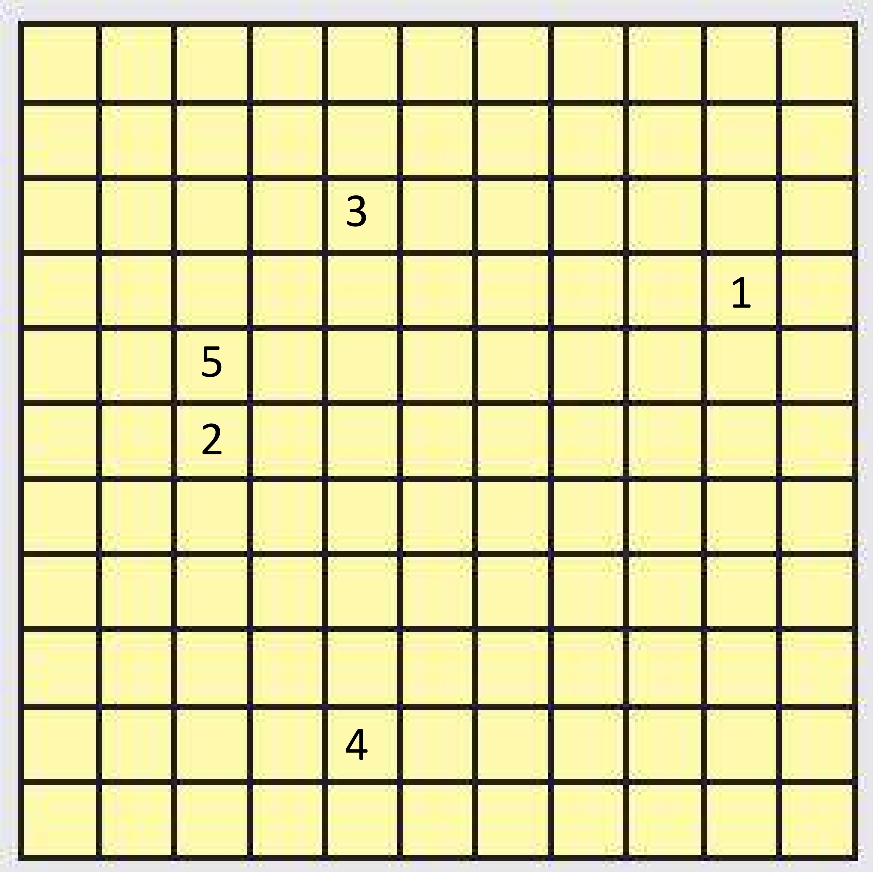
- scanning the training images with a fixed template to extract a list of patches



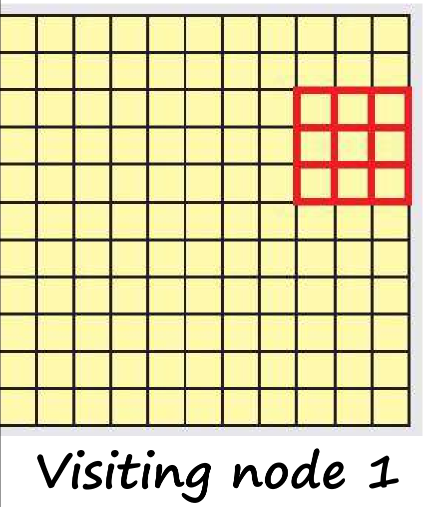
Arpat, G. B. (2004). SIMPAT: Stochastic Simulation with Patterns.

Basic idea

- define a random path on the simulation grid
    - Empty simulation grid on which a realization is generated by sequentially visiting each node
    - Shows first 5 visiting nodes

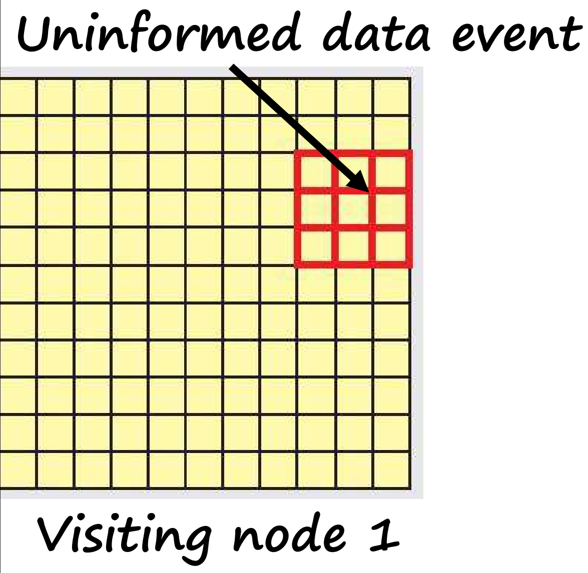


Basic idea

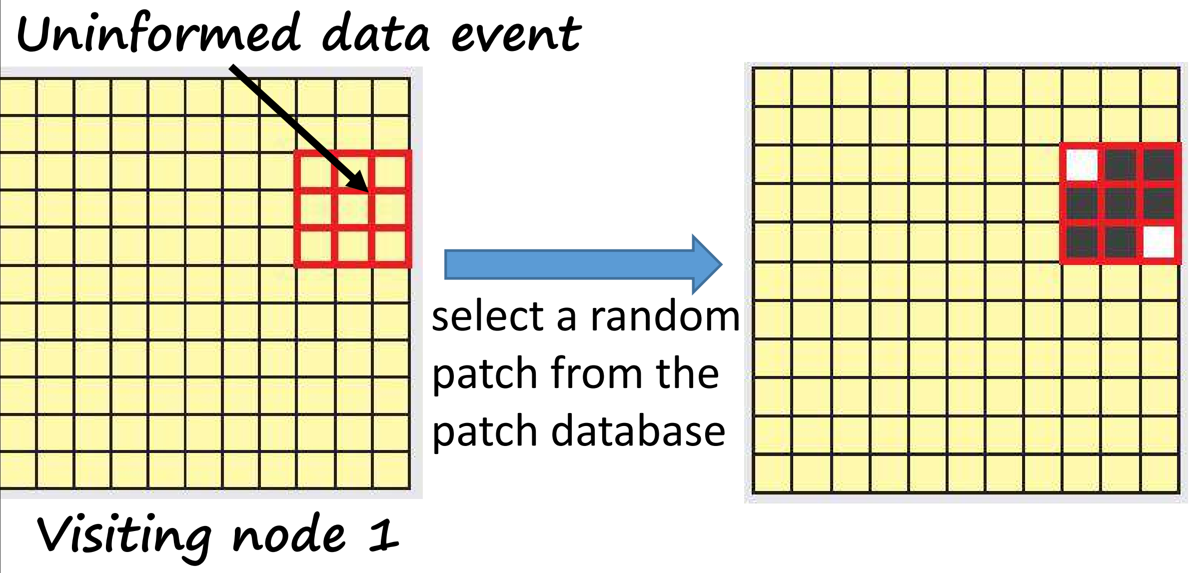


Arpat, G. B. (2004). SIMPAT: Stochastic Simulation with Patterns.

Basic idea



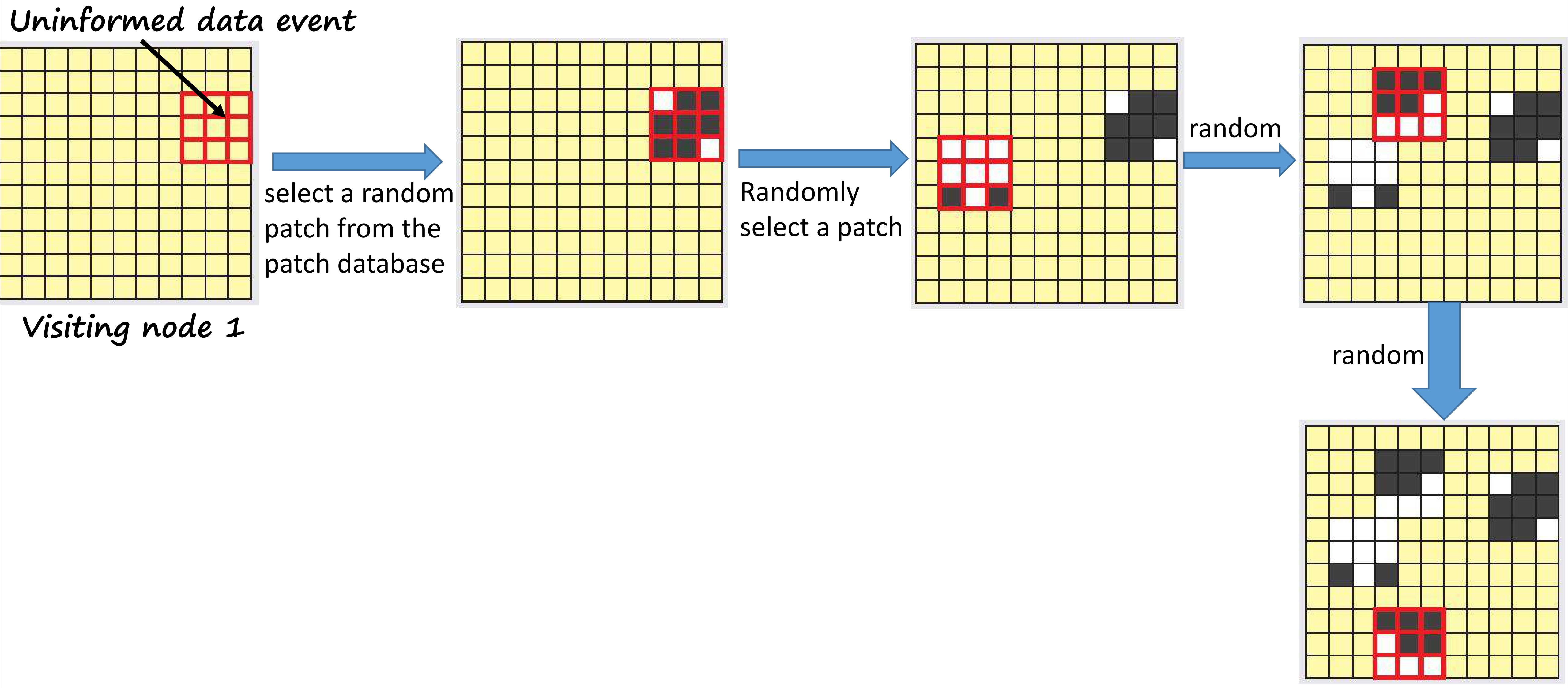
Basic idea



compare

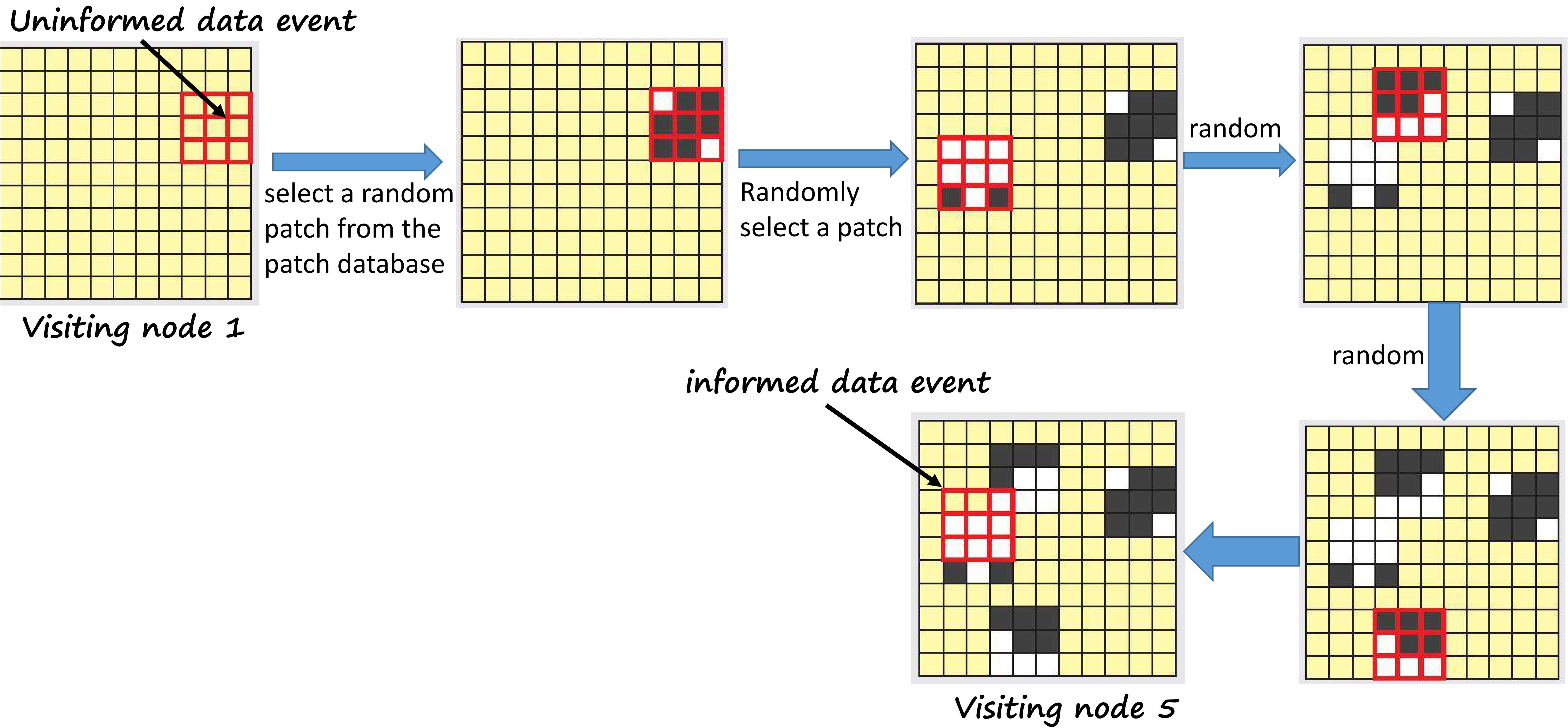
Arpat, G. B. (2004). SIMPAT: Stochastic Simulation with Patterns.

Basic idea



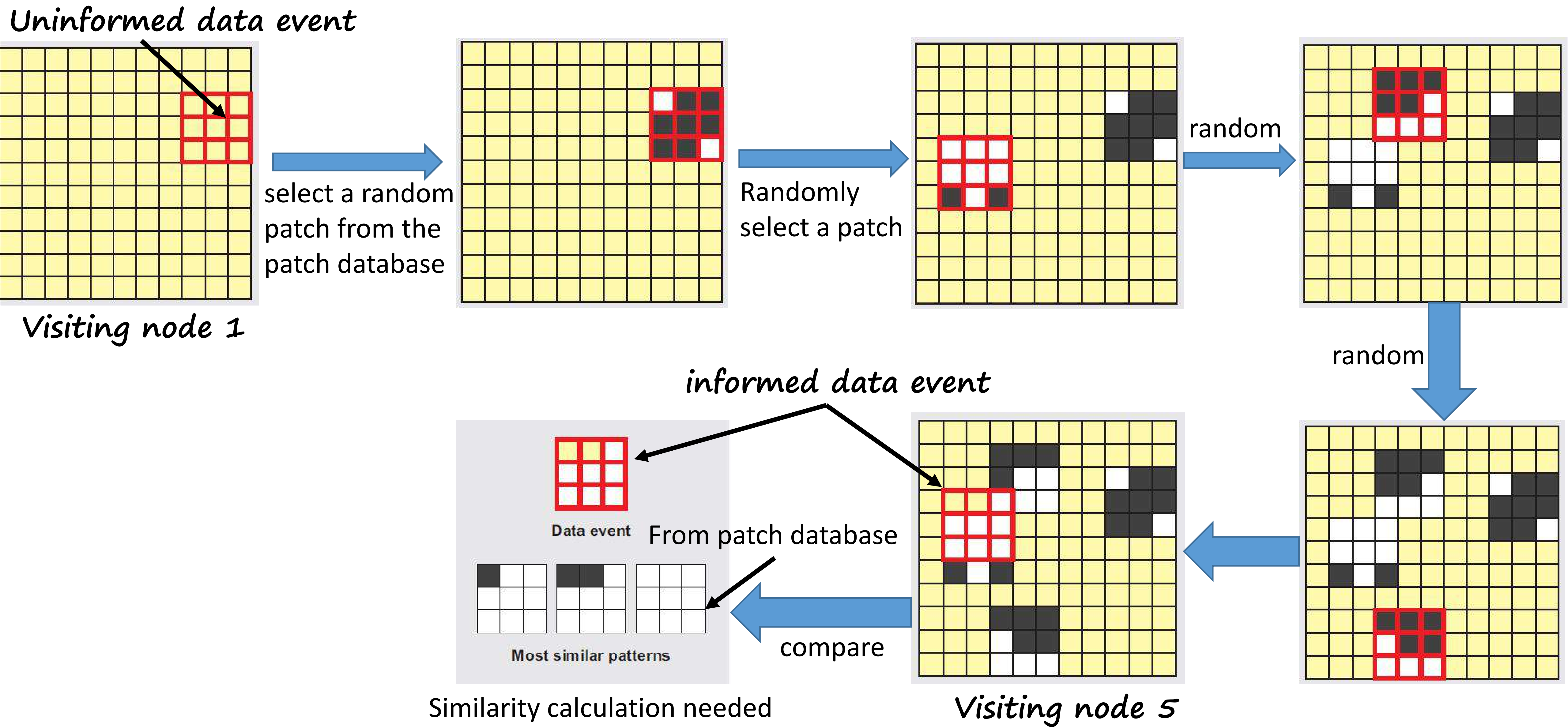
Arpat, G. B. (2004). SIMPAT: Stochastic Simulation with Patterns.

Basic idea



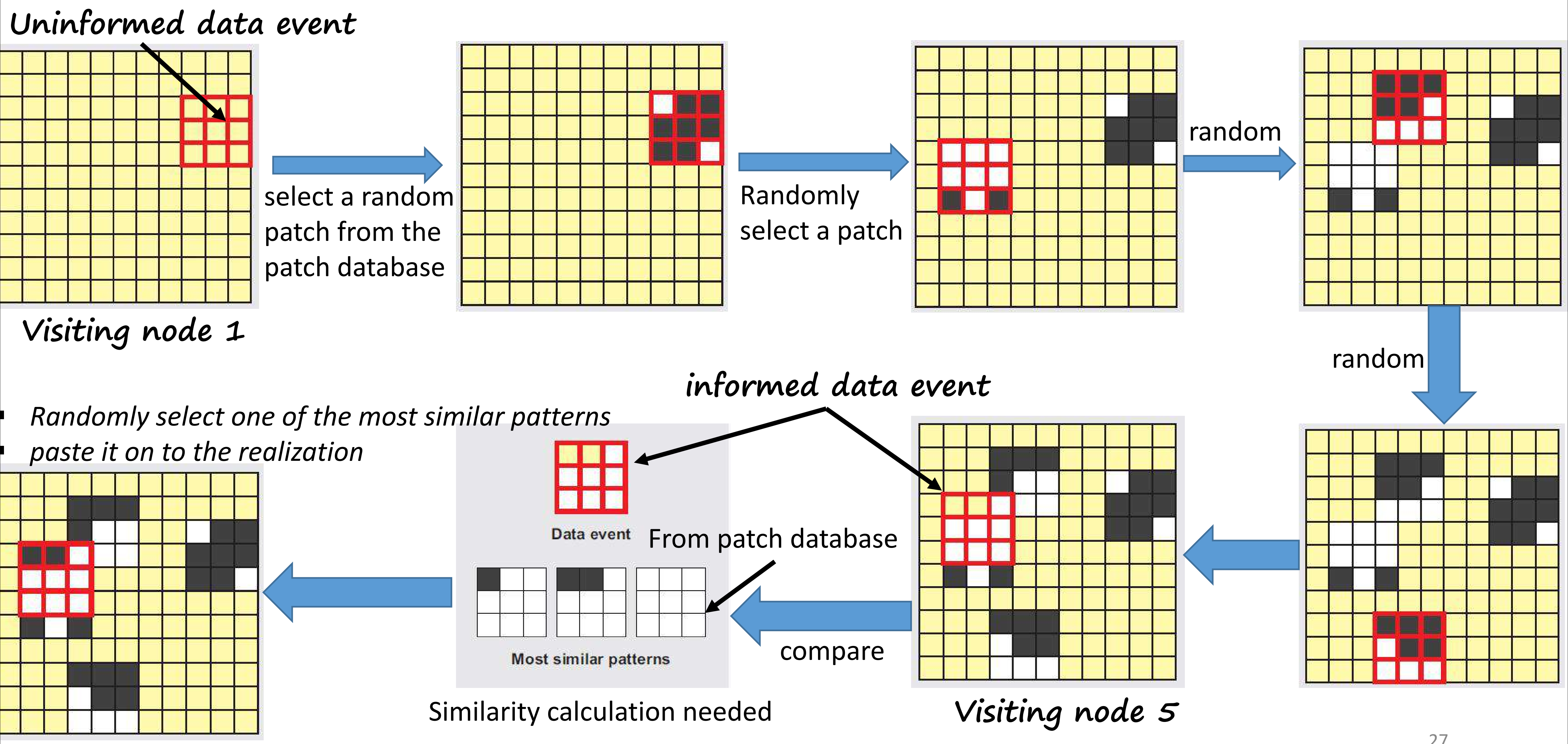
Arpat, G. B. (2004). SIMPAT: Stochastic Simulation with Patterns.

Basic idea



Arpat, G. B. (2004). SIMPAT: Stochastic Simulation with Patterns.

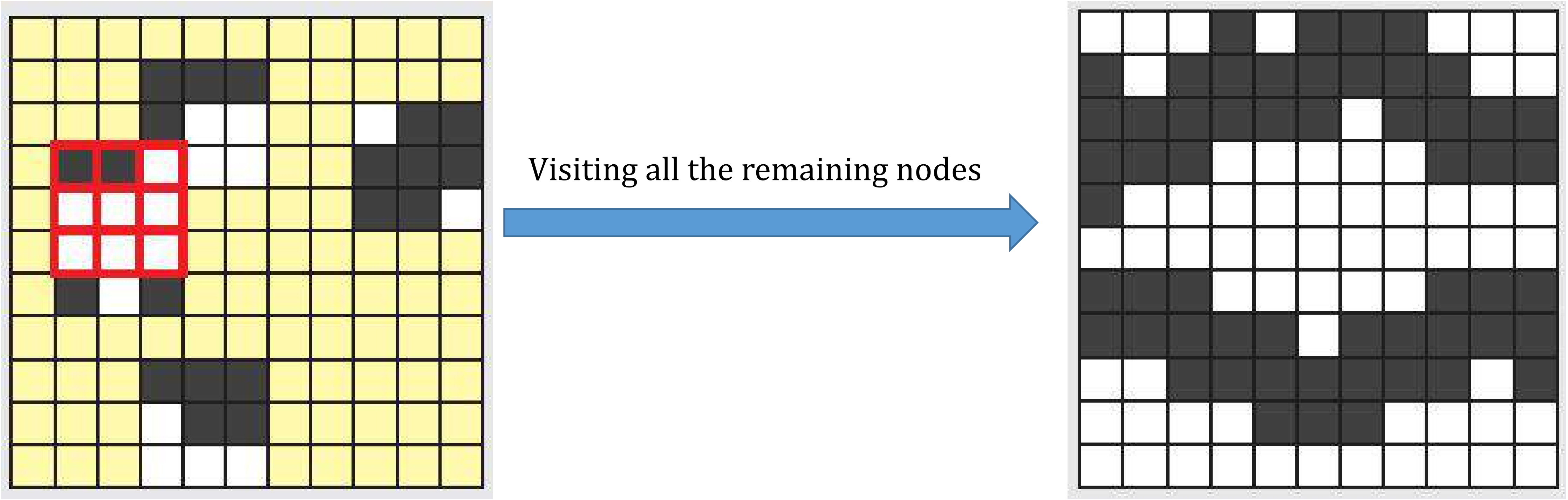
Basic idea



Arpat, G. B. (2004). SIMPAT: Stochastic Simulation with Patterns.

Basic idea

- Visiting all the remaining nodes in this fashion, a final realization is obtained



Final realization

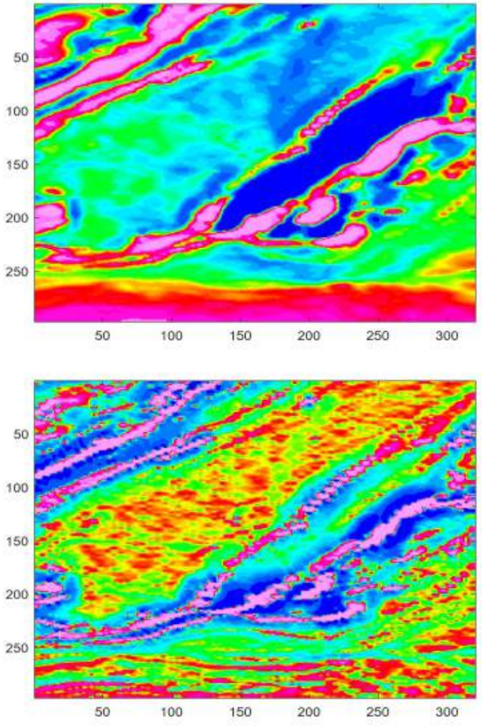
Arpat, G. B. (2004). SIMPAT: Stochastic Simulation with Patterns.

Notes

- Similarity calculation
    - Use concept of distance
- Many different realizations can be generated by changing some parameters in MPS
    - Order of visiting nodes (random seed)
    - Size of the template
- The idea of MPS used in data augmentation is much more complex than the basic idea shown above
    - Multivariate TI
    - Nonstationary TI:
        - Geological patterns tend to vary in space,
        - statistically different over the entire spatial domain

Traditional Multi-point geostatistics

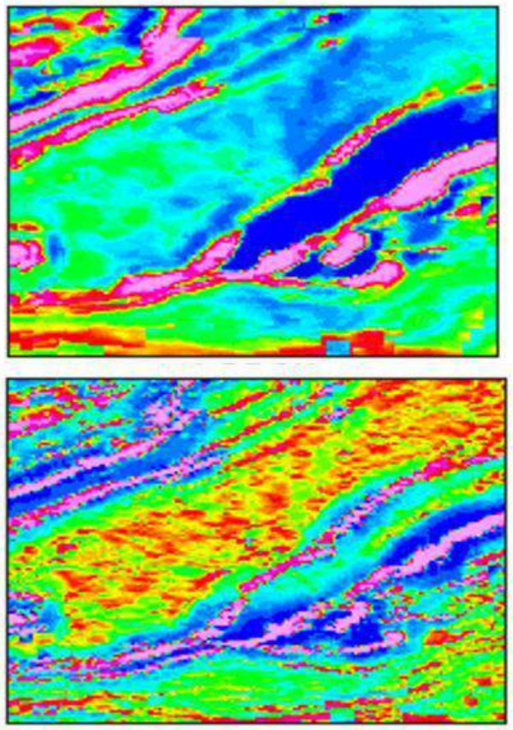
Training image



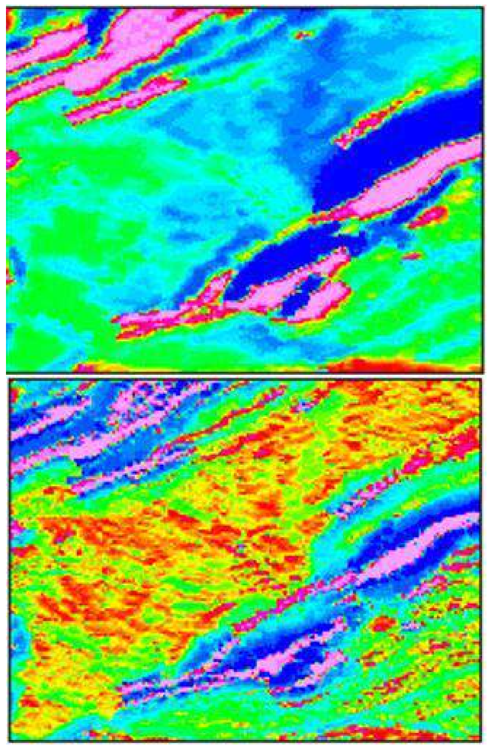
Multiple-point statistics

- Geologically feasible
- Same spatial continuity

Realization 1



Realization 2

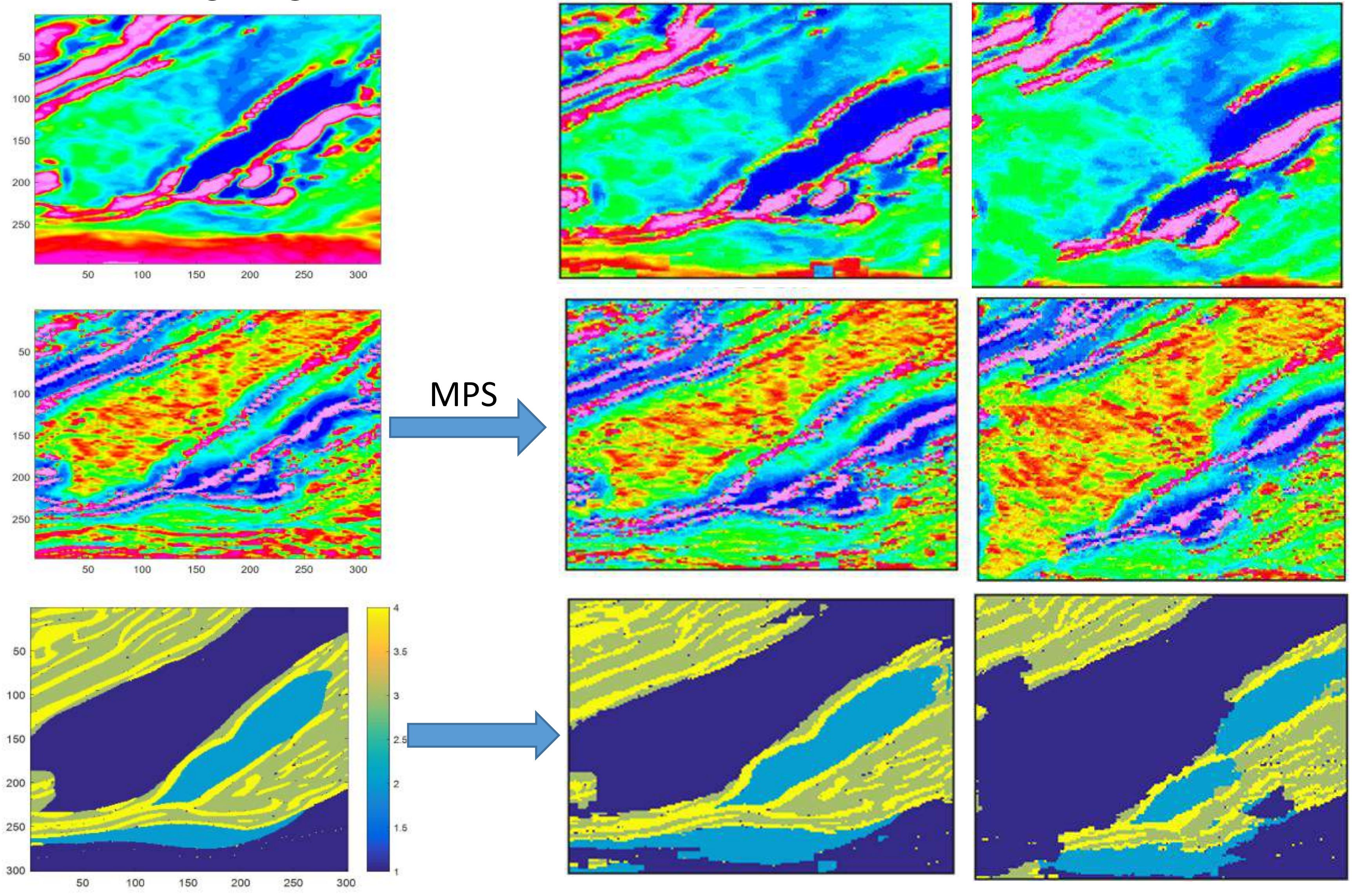


MPS → Data augmentation

- Different from traditional MPS
  - ➢Not only simulate geophysical images,
  - ➢but also simulate corresponding labels

Data augmentation

Training image | Realization 1 | Realization 2



- Tune parameters
  - Order of visiting nodes
  - Size of template
  - Etc.

- ❑ generate different realizations
- ❑ Realization 1 more similar to TI

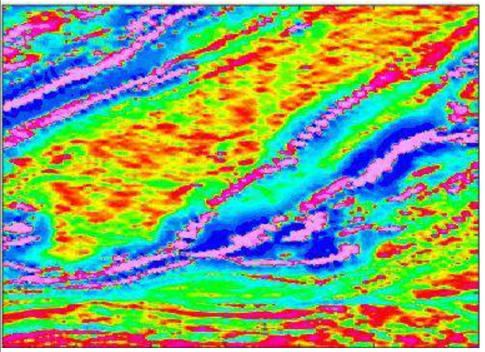
Methods

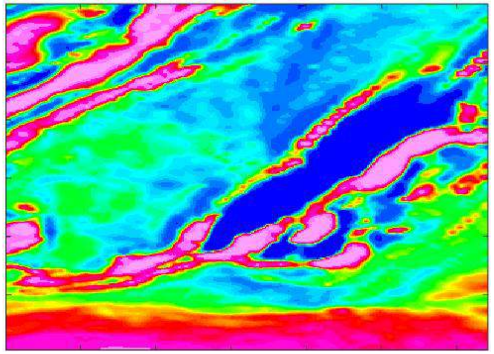
- Data augmentation
- Supervised learning using iCNN

iCNN

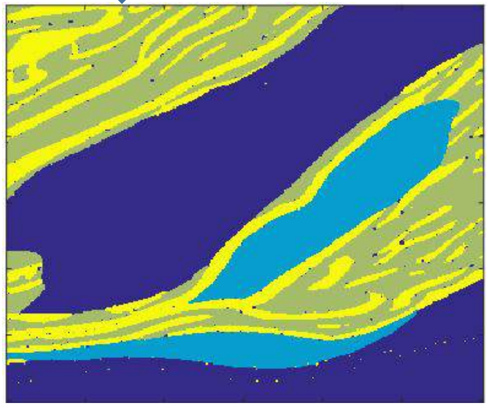
Why **new NN architecture**?

- more complex image-label relationship
- tried resnet , good for cat segmentation, but not good for geological segmentation

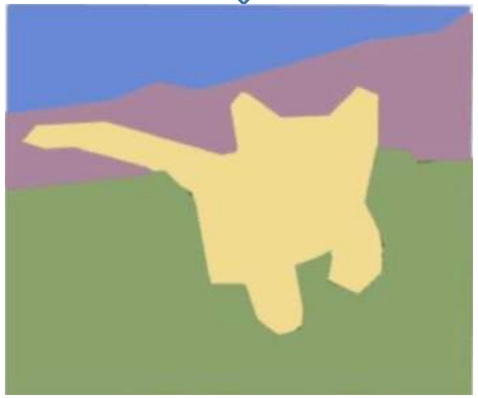




- Experienced geologists
- Done subjectively



- Any schoolchild quite likely to know how to segment



Resnet vs. iCNN

- Mathematical Viewpoint

- Architectural Viewpoint

Resnet vs. iCNN

- *Resnet*
  - $Y_{j+1} = Y_j + \sigma\left(N\left(K_j Y_j\right) + b_j\right) \qquad j = 0,\ 1, \cdots, N-1$
- *iCNN*
  - $Y_{j+\frac{1}{2}} = Y_j + h\sigma(N\left(S_j Y_j\right) + b_j)$
  - $Y_{j+1} = (I + hK_j^T K_j)^{-1} Y_{j+\frac{1}{2}}$

Resnet vs. iCNN

- *Resnet*

  - $Y_{j+1} = Y_j + \sigma\left(N\left(K_j Y_j\right) + b_j\right) \qquad j = 0,\ 1, \cdots, N-1$

- *iCNN*

  - $Y_{j+\frac{1}{2}} = Y_j + h\sigma(N\left(S_j Y_j\right) + b_j)$

  - $Y_{j+1} = (I + hK_j^T K_j)^{-1} Y_{j+\frac{1}{2}}$

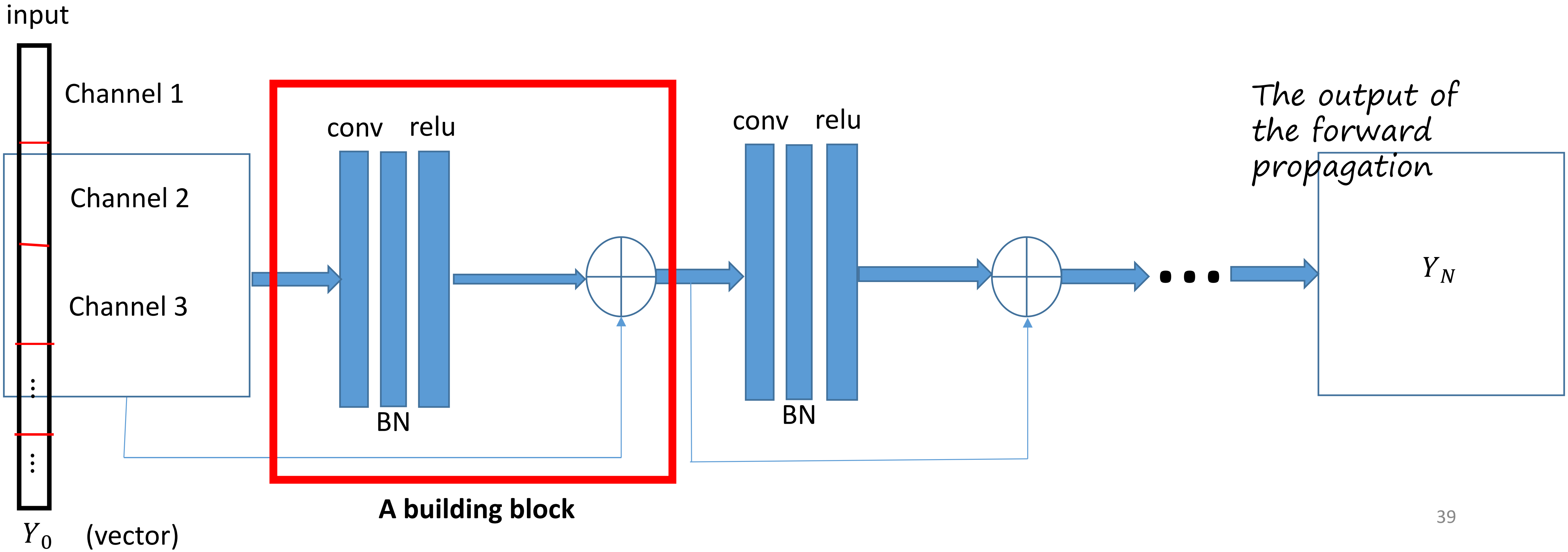
❑ additional step
❑ involve inverse matrix

Resnet vs. iCNN

- Mathematical Viewpoint
- Architectural Viewpoint

A typical building block of Resnet

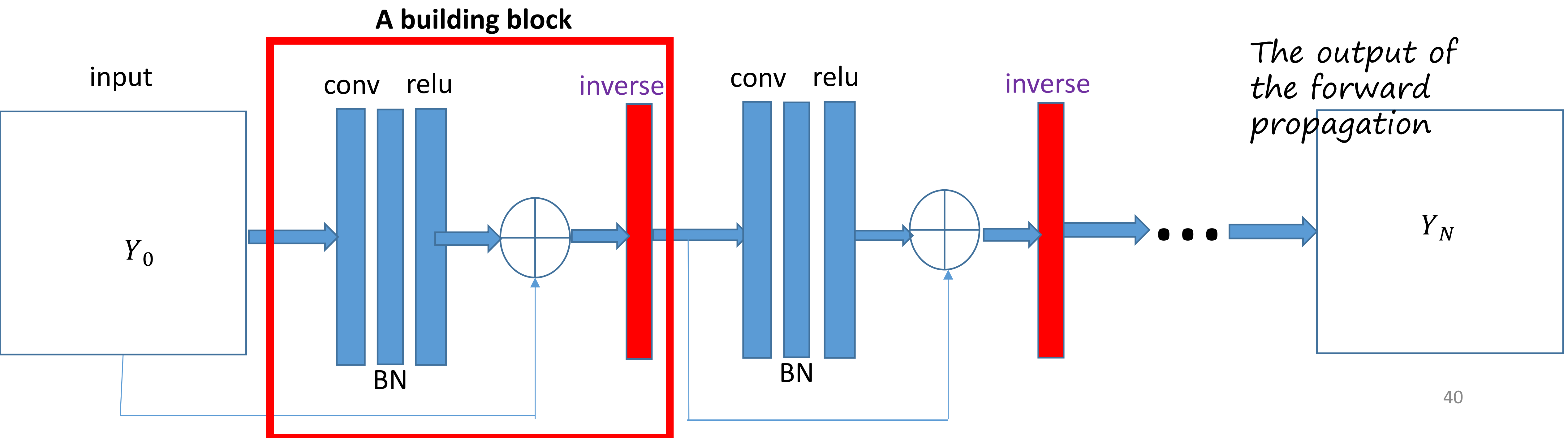
- *nonlinearly transform the input features*
- $Y_{j+1} = Y_j + \sigma\left(N\left(K_j Y_j\right) + b_j\right) \qquad j = 0,\ 1, \cdots, N-1$

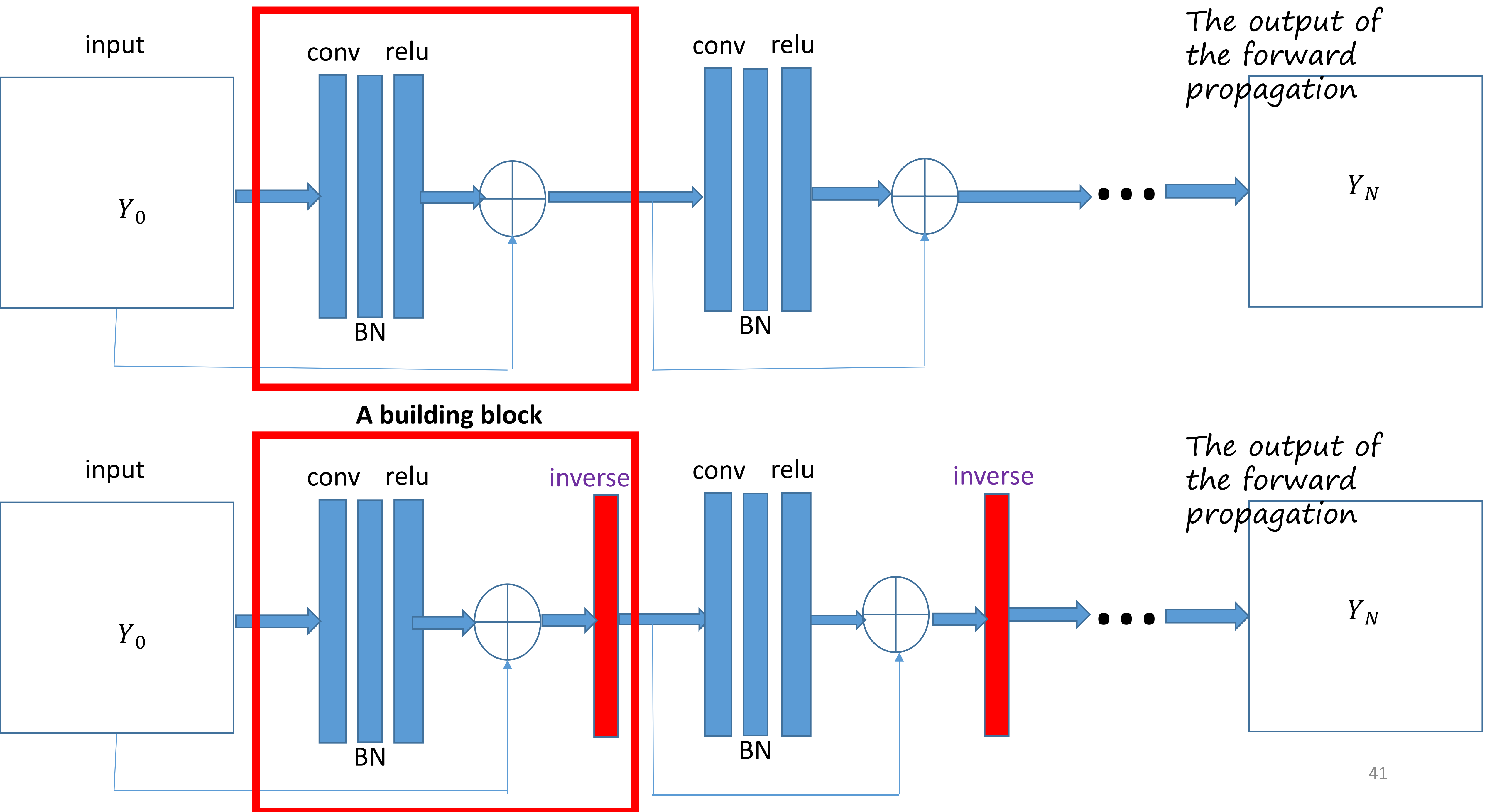


A building block

A typical building block of iCNN

- *nonlinearly transform the input features*
- $Y_{j+\frac{1}{2}} = Y_j + h\sigma(N(S_jY_j) + b_j)$
- $Y_{j+1} = (I + hK_j^TK_j)^{-1}Y_{j+\frac{1}{2}}$



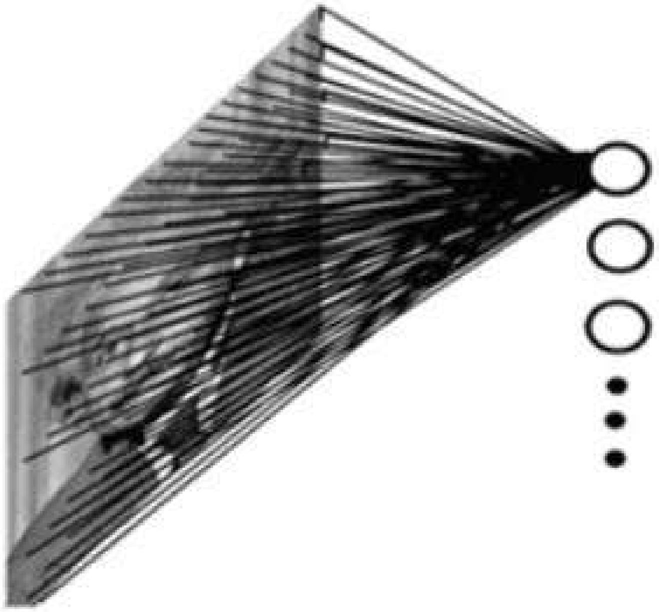

input
Y_0
conv
relu
BN
conv
relu
BN
The output of the forward propagation
Y_N
A building block
input
Y_0
conv
relu
BN
inverse
conv
relu
BN
inverse
The output of the forward propagation
Y_N

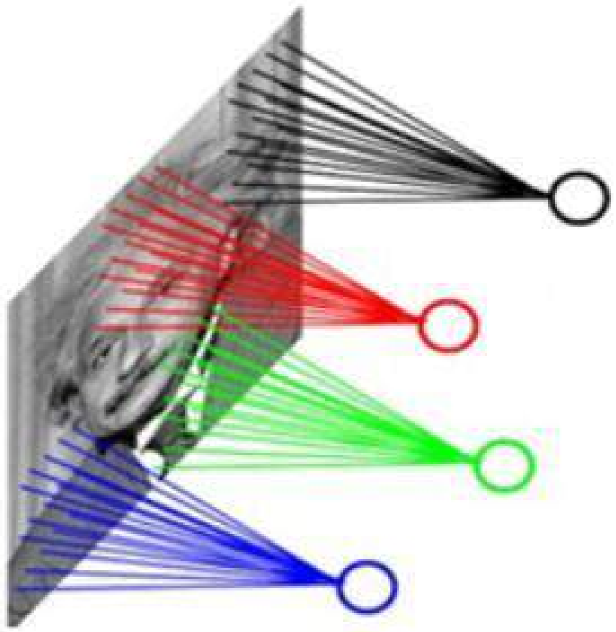
iCNN

Question addressed:

Why iCNN, not Resnet?

iCNN	Resnet
Inverse of convolution matrix	Convolution matrix



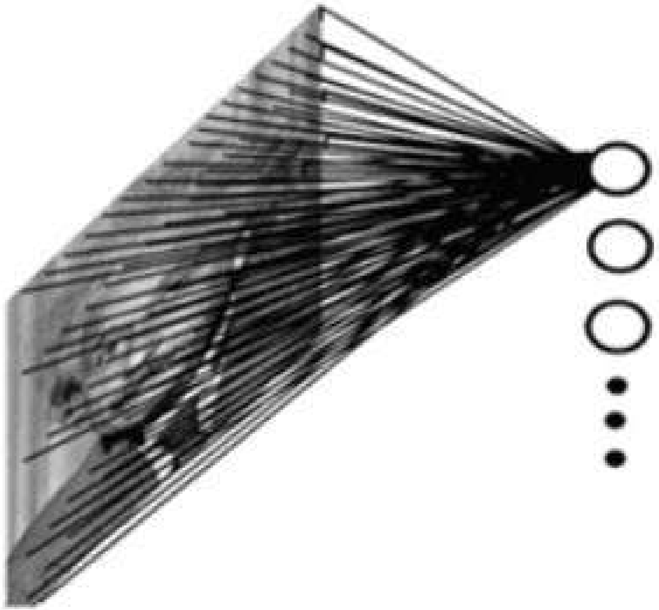


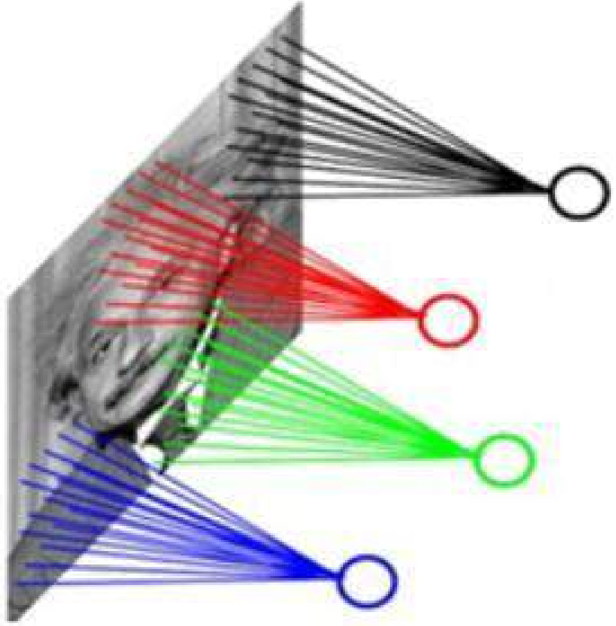
iCNN

Question addressed:

Why iCNN, not Resnet?

iCNN	Resnet
Inverse of convolution matrix	Convolution matrix
Large dense matrix	Large sparse matrix



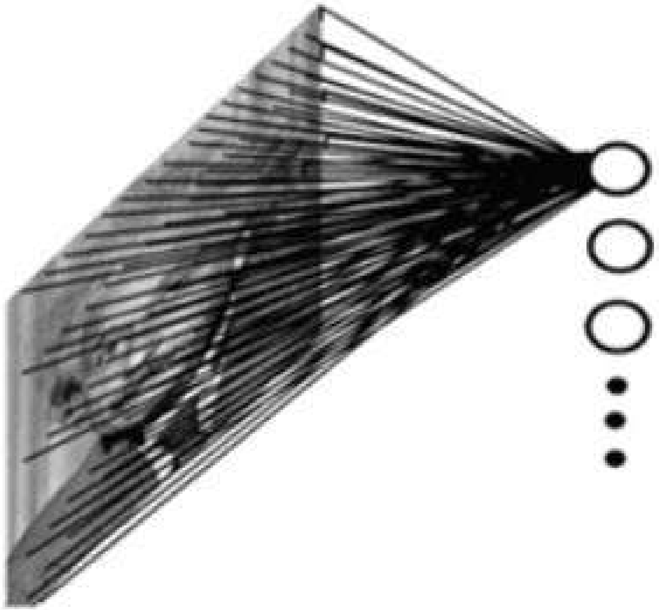


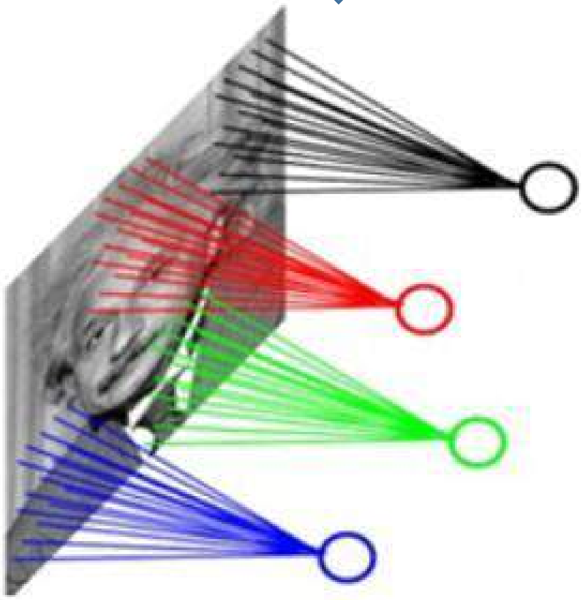
iCNN

Question addressed:

Why iCNN, not Resnet?

iCNN	Resnet
Inverse of convolution matrix	Convolution matrix
Large dense matrix	Large sparse matrix
Globally connected	Locally connected





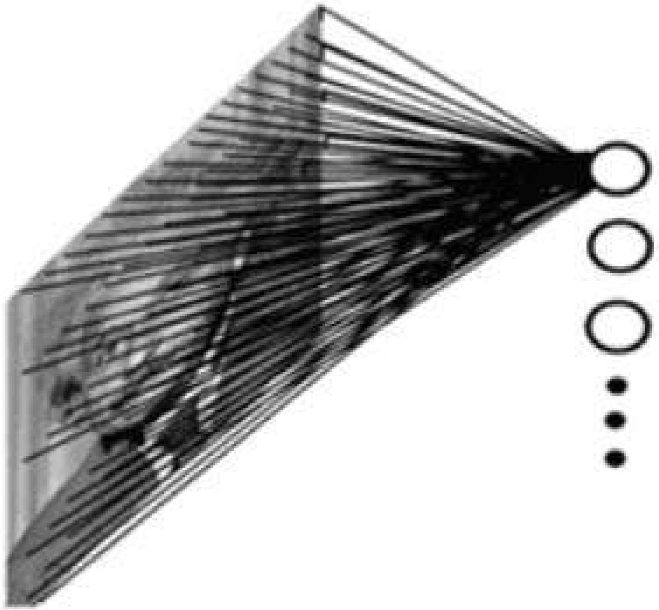
iCNN

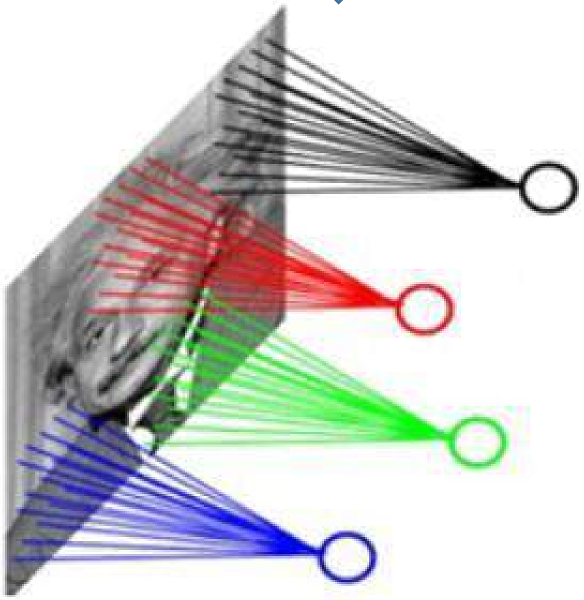
Question addressed:

Why iCNN, not Resnet?

iCNN	Resnet
Inverse of convolution matrix	Convolution matrix
Large dense matrix	Large sparse matrix
Global connection	Local connection

- Good for different scales contained in geological data





iCNN

Question addressed:

How to derive it?

- $Y_{j+1} = Y_j + f(Y_j; \theta_j)$ start from Resnet
    - $f(Y; \theta) = f(Y; W, b) = \sigma(KY + b)$ activity function
- $Y_{j+1} = Y_j + hf(Y_j; \theta_j)$ consider as a continuous process
- $\frac{Y_{j+1} - Y_j}{h} = f(Y_j; \theta_j)$ an explicit Euler discretization of the ODE
- $\dot{y}(t) = f(y(t); \theta(t))$ a fully continuous process

iCNN

How to derive it?

Start from Resnet: $Y_{j+1} = Y_j + f(Y_j; \theta_j)$

Consider Resnet as a continuous process

$$Y_{j+1} = Y_j + hf(Y_j; \theta_j)$$

$$\dot{y}(t) = f(y(t); \theta(t))$$

The diffusion reaction network

$$\dot{y}(t) = -K(t)^T K(t) y(t) + \sigma(N(\boldsymbol{S(t)}Y) + b(t))$$

- *a diagonal convolution operator*
- 1D convolution

IMEX discretization $Y_{j+\frac{1}{2}} = Y_j + h\sigma\left(N\left(S_j Y_j\right) + b_j\right)$

$$Y_{j+1} = (I + hK_j^T K_j)^{-1} Y_{j+\frac{1}{2}}$$

$iCNN$

- $\dot{y}(t) = f\big(y(t); \theta(t)\big)$

iCNN

- $\dot{y}(t) = f\big(y(t); \theta(t)\big)$
- $\dot{y}(t) = -K(t)^T K(t) y(t) + f\big(y(t); \theta(t)\big)$

- Diffusion term
  - ❑ Stable
  - ❑ Diffuse, Globally connected

iCNN

- $\dot{y}(t) = f\big(y(t); \theta(t)\big)$

- Diffusion term
  - ❑ Stable
  - ❑ Diffuse, Globally connected

- $\dot{y}(t) = -K(t)^T K(t) y(t) + f\big(y(t); \theta(t)\big)$
  - $K(t) = \begin{pmatrix} K_1(t) & & \\ & \ddots & \\ & & K_m(t) \end{pmatrix}$

❑ a diagonal convolution operator
❑ Imitate vector heat equation
❑ $y(t)$ has m channels for m different kernels

iCNN

- $\dot{y}(t) = f\big(y(t); \theta(t)\big)$
- $\dot{y}(t) = -K(t)^T K(t) y(t) + f\big(y(t); \theta(t)\big)$

Reaction term

iCNN

- $\dot{y}(t) = f\big(y(t); \theta(t)\big)$
- $\dot{y}(t) = -K(t)^T K(t) y(t) + f\big(y(t); \theta(t)\big)$
  - $f\big(y(t); \theta(t)\big) = \sigma(N(\boldsymbol{S}(\boldsymbol{t})Y) + b(t))$

Reaction term

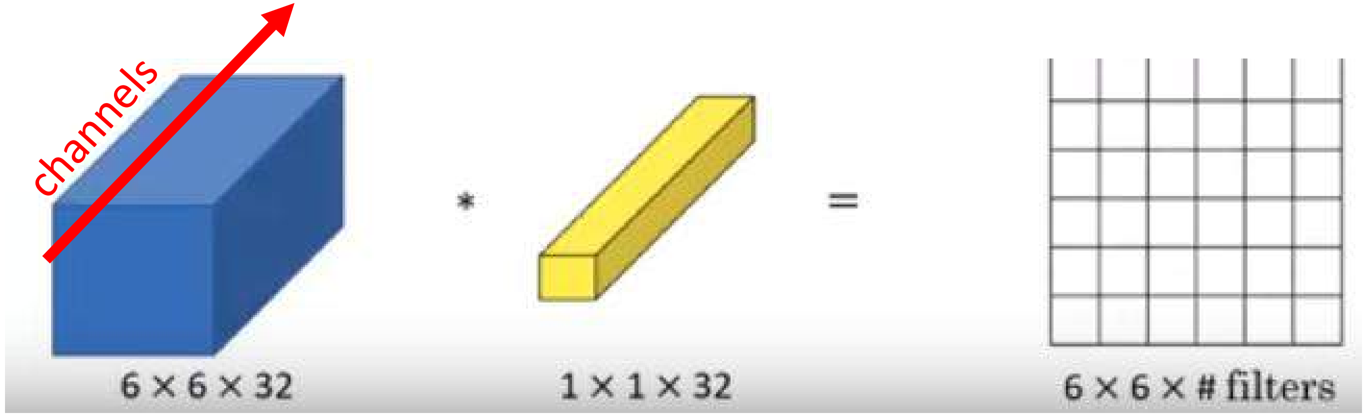
- 1D convolution
  - ❑ couples channels
  - ❑ not couple any pixels in space

iCNN

- $\dot{y}(t) = f\big(y(t); \theta(t)\big)$
- $\dot{y}(t) = -K(t)^T K(t) y(t) + f\big(y(t); \theta(t)\big)$
    - $f\big(y(t); \theta(t)\big) = \sigma(N(\boldsymbol{S}(\boldsymbol{t})Y) + b(t))$

Reaction term

- 1D convolution
    - ❑ couples channels
    - ❑ not couple any pixels in space



Y S

IMEX

- *linear term*
- *decay fast*

- *nonlinear term*
- *decay slow*

- $\dot{y}(t) = -K(t)^T K(t) y(t) + f\big(y(t); \theta(t)\big)$
  - *Discretization*
    - Implicit Explicit schemes (IMEX)

- $Y_{j+\frac{1}{2}} = Y_j + h\sigma\big(N(S_j Y_j) + b_j\big)$
- $Y_{j+1} = (I + hK_j^T K_j)^{-1} Y_{j+\frac{1}{2}}$

- *nonlinear term*
  - Explicit step
  - Require small h to be stable
- *linear term*
  - Implicit step
    - Non-local
    - Unconditional stable

Computation

- $(I + hK_j^T K_j)^{-1}$
  - Involve inverse of a matrix
  - Impose periodic boundary condition on K_j
    - Block Circulant with Circulant Blocks (BCCB)
    - Decomposed as $K_j = F^* \Lambda_i F$

2D unitary (DFT) matrix

diagonal matrix containing the eigenvalues of K_j

Computation

- $K_j = F^* \Lambda_i F$

- $K_j^T K_j = F^* \left(\Lambda_j^* \Lambda_j\right) F$

Point-wise division

- $(hK^T K + I)^{-1} = \begin{Bmatrix} F^* \left(\frac{1}{1+h\Lambda_1^* \Lambda_1}\right) F & & \\ & \ddots & \\ & & F^* \left(\frac{1}{1+h\Lambda_1^* \Lambda_1}\right) F \end{Bmatrix}$

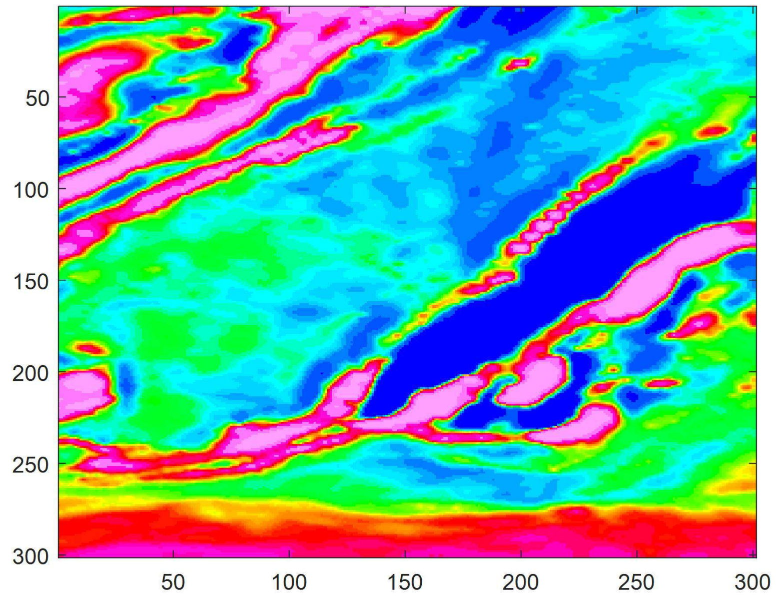
- $(hK^T K + I)^{-1} Y \quad easy!$

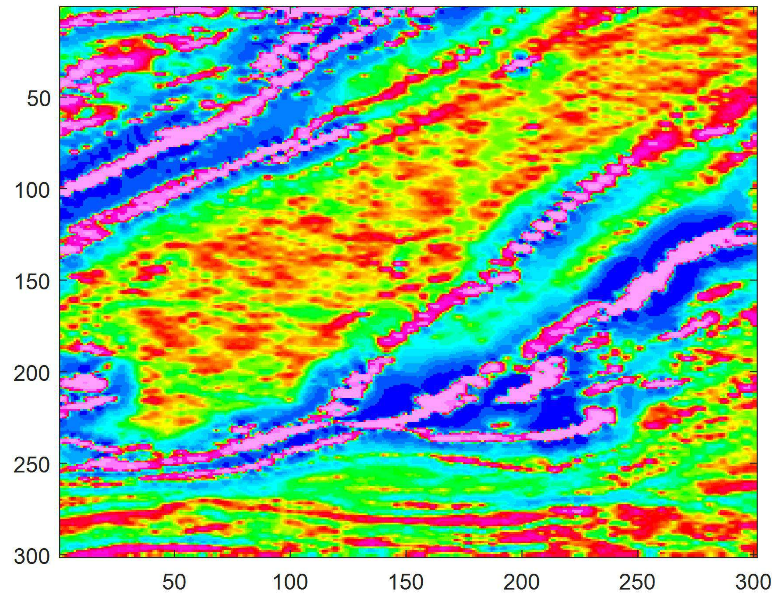
Computation

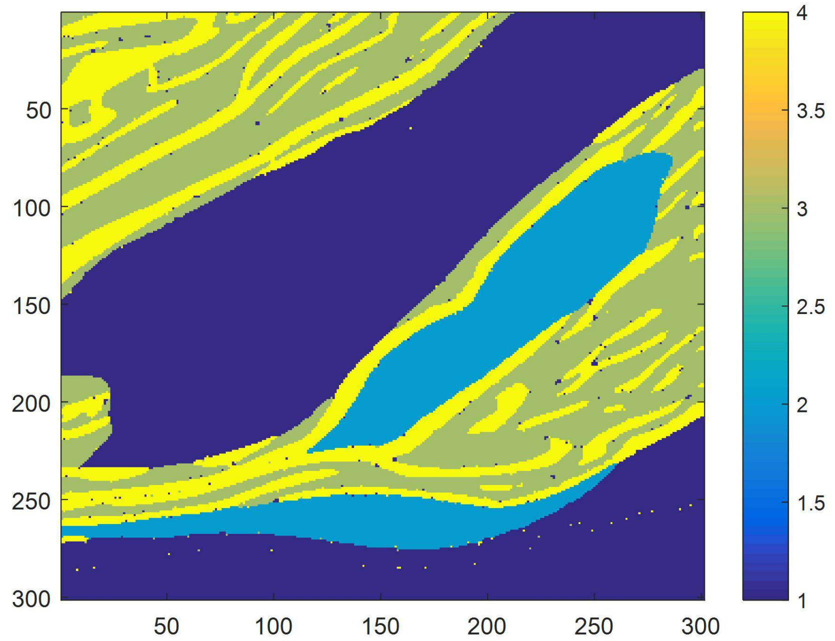
- $(hK^TK + I)^{-1}Y = \begin{Bmatrix} F^*\left(\frac{1}{1+h\Lambda_1^*\Lambda_1}\right)F & & \\ & \ddots & \\ & & F^*\left(\frac{1}{1+h\Lambda_1^*\Lambda_1}\right)F \end{Bmatrix} Y$

- Compute the FFT of the different channels of Y
- Compute the FFT of the convolution matrix K (FK = ΛF)
- Pointwise divide each channel by $1 + h\Lambda_1^*\Lambda_1$
- Transform the result using the inverse FFT transform

Results



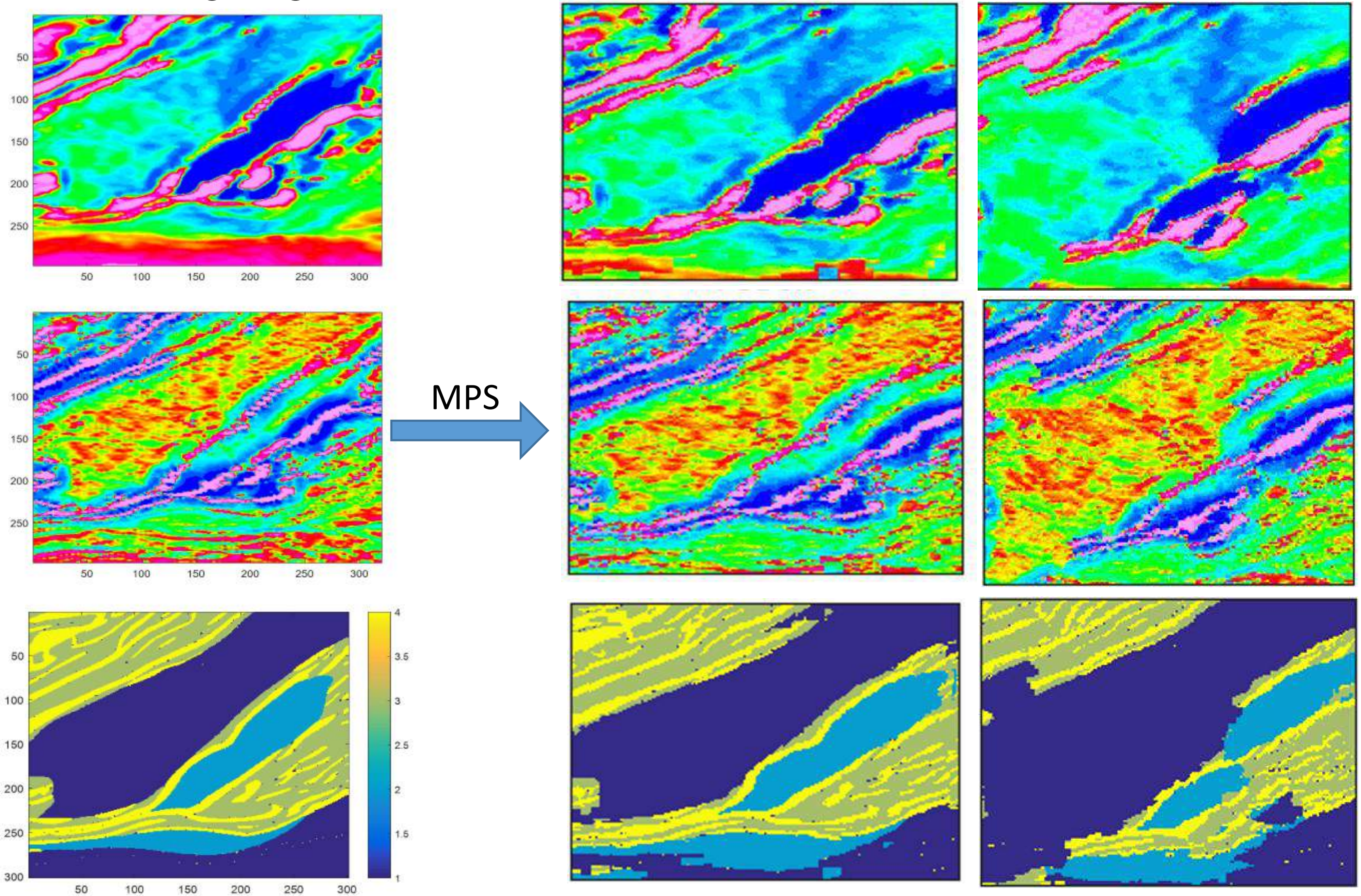




- *Geological unit map*
  - *provided by an expert geologist*

Data augmentation

Training image | Realization 1 | Realization 2

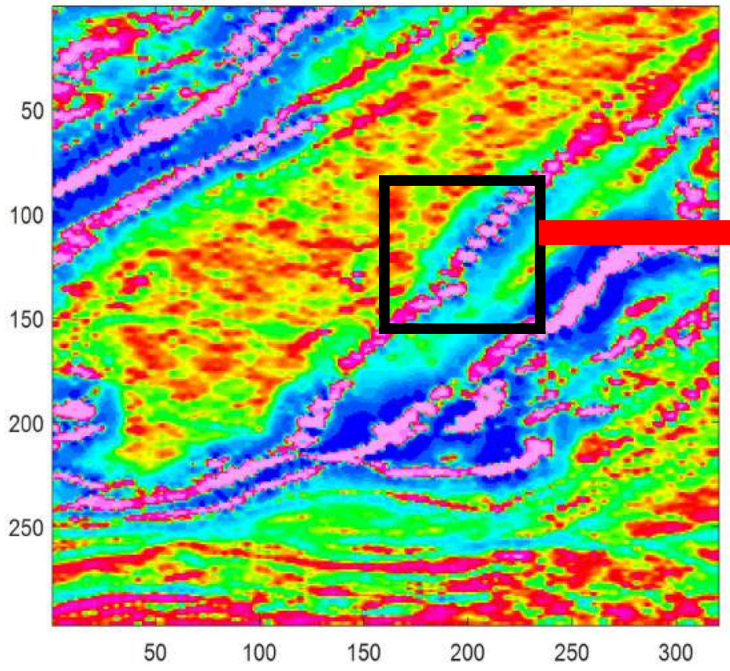


- Geologically feasible
- Same spatial continuity

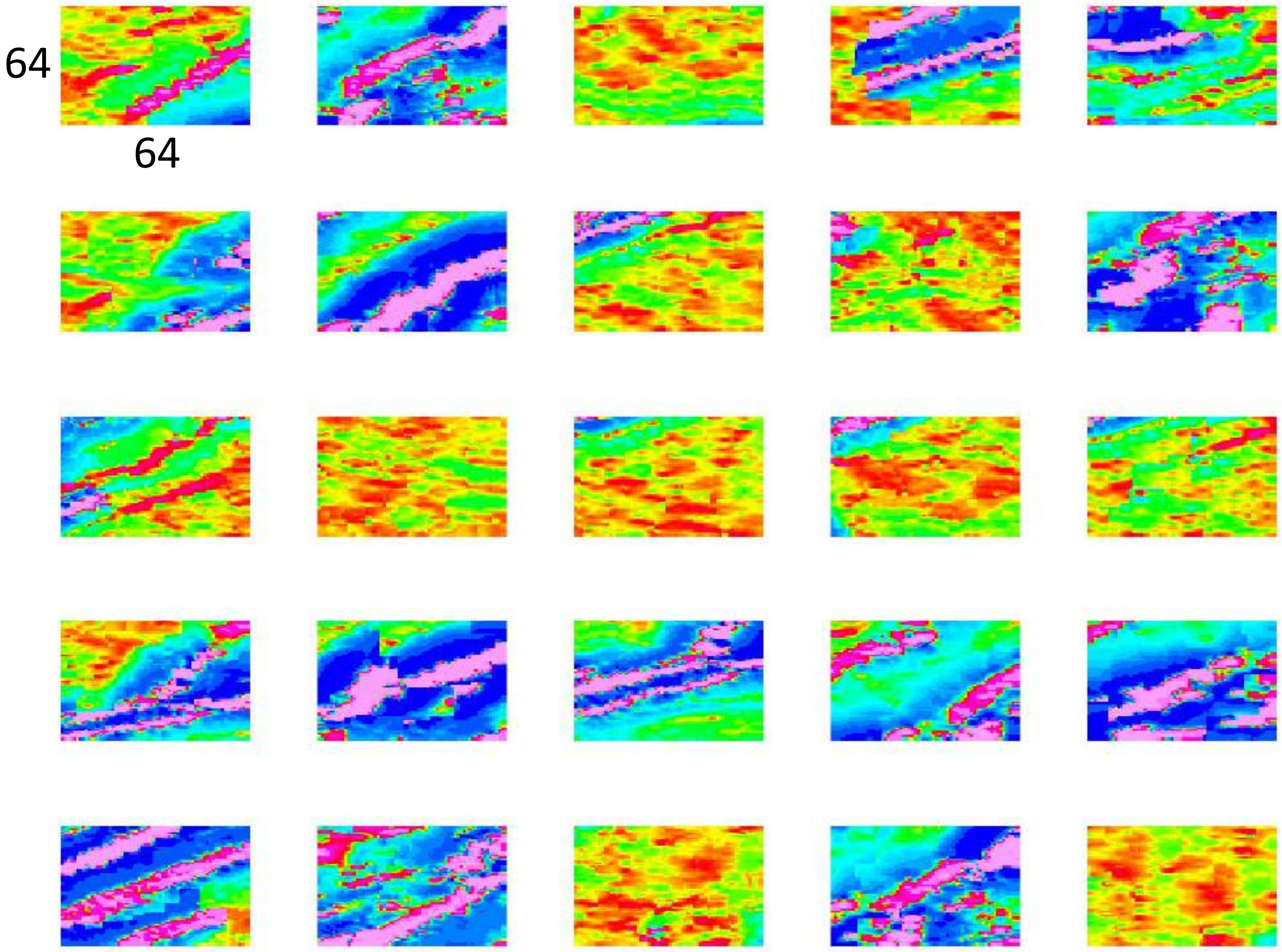
- Tune parameters
    - Order of visiting nodes
    - Size of template
    - Etc.

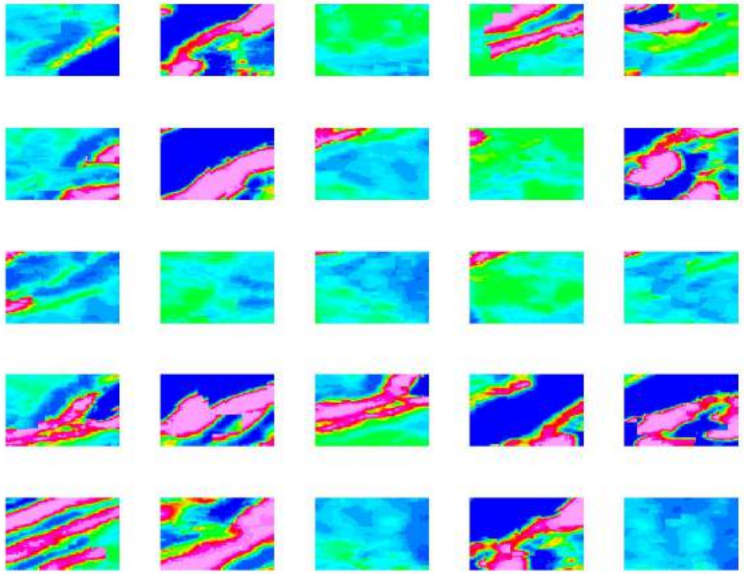
- ❑ generate different realizations
- ❑ Realization 1 more similar to TI

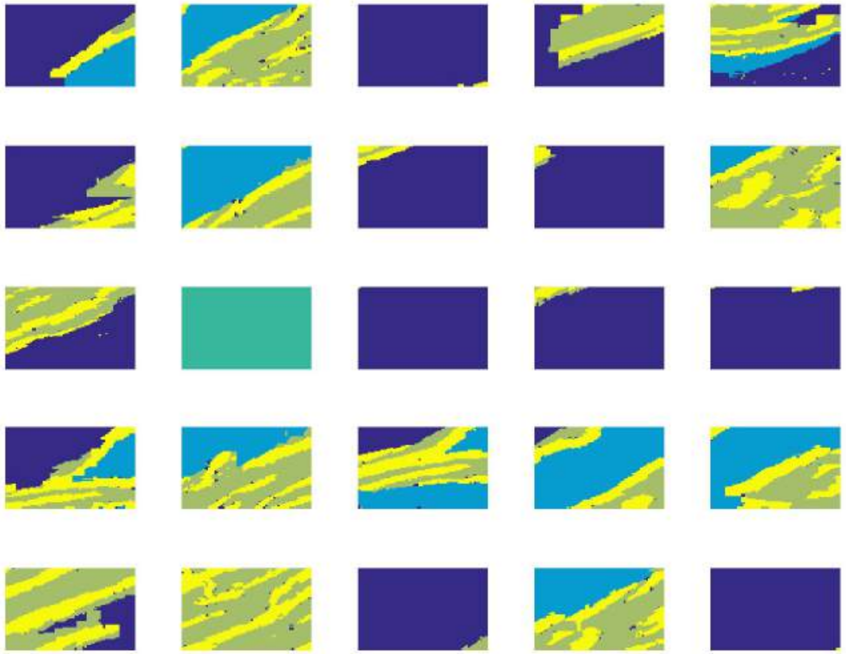
Randomly cropping



Training examples

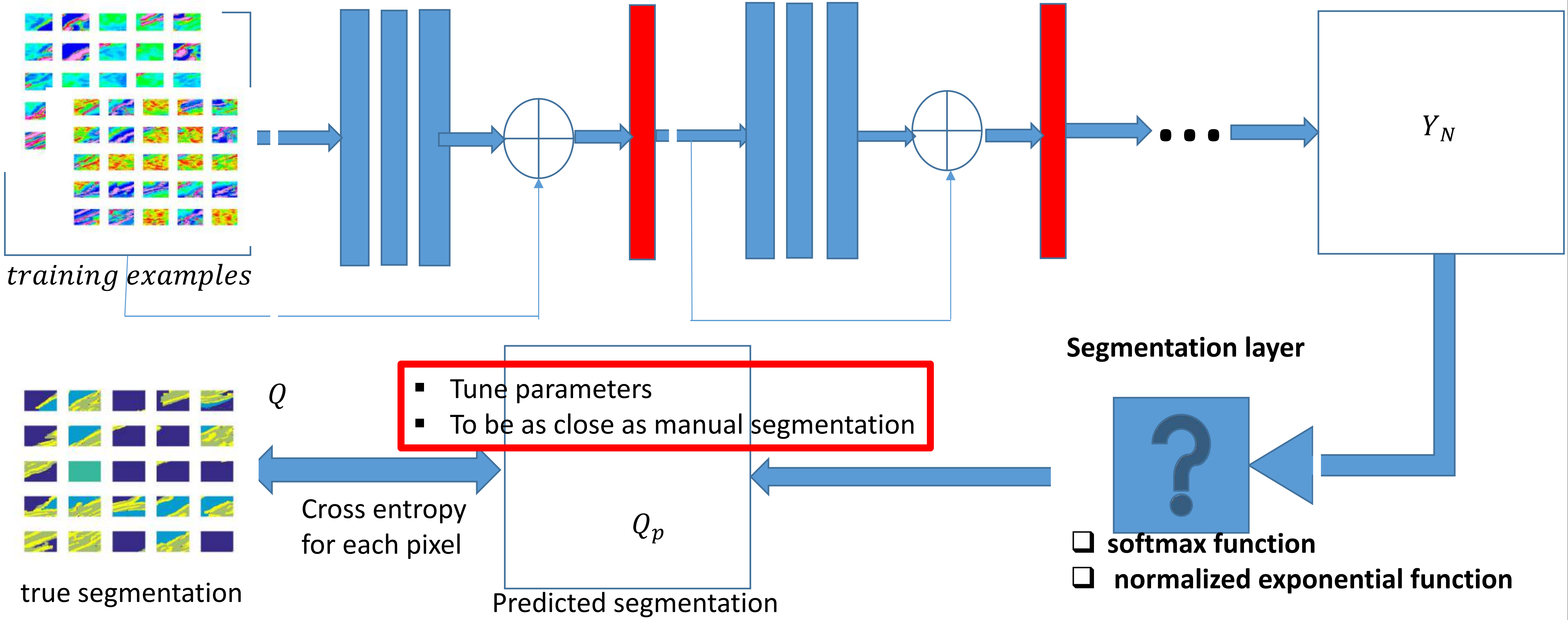




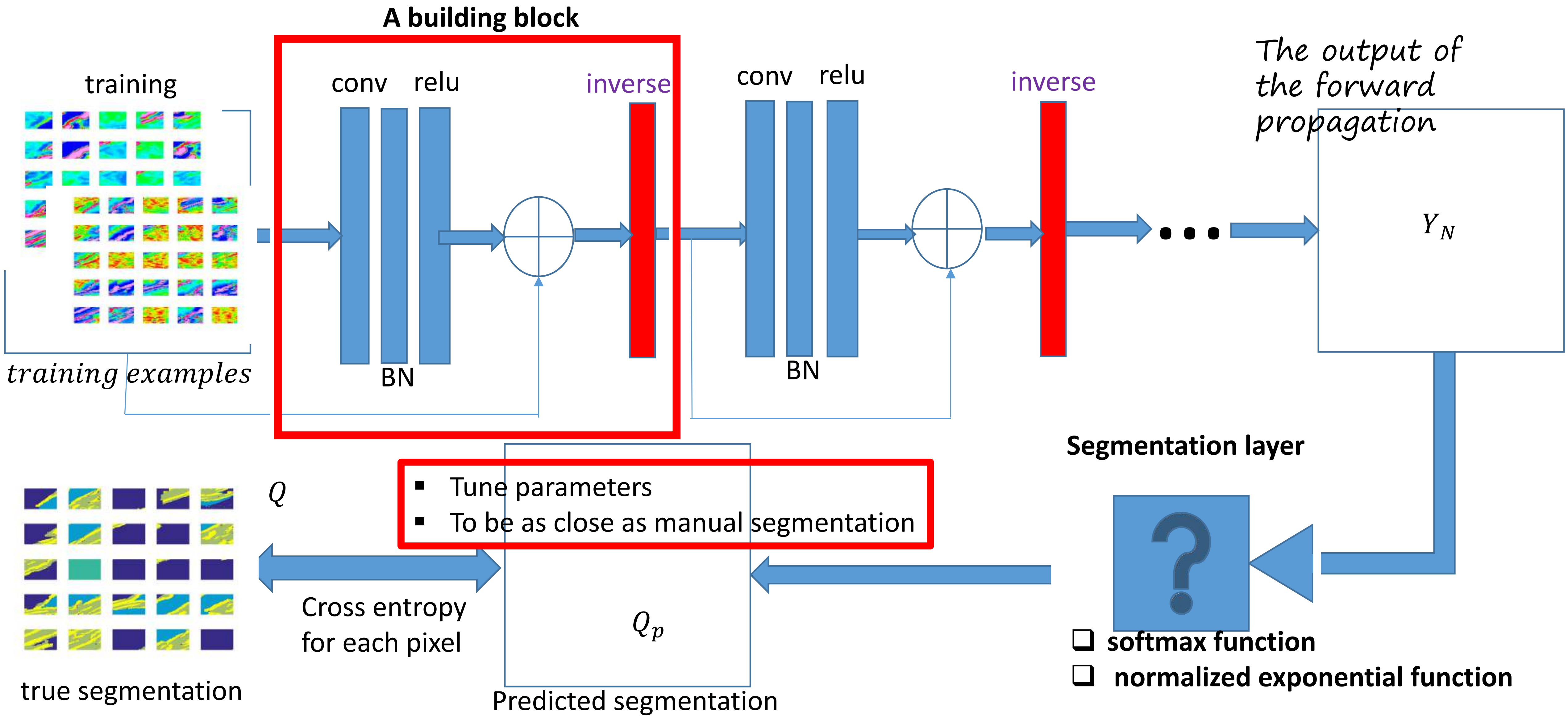


Training dataset (3000)

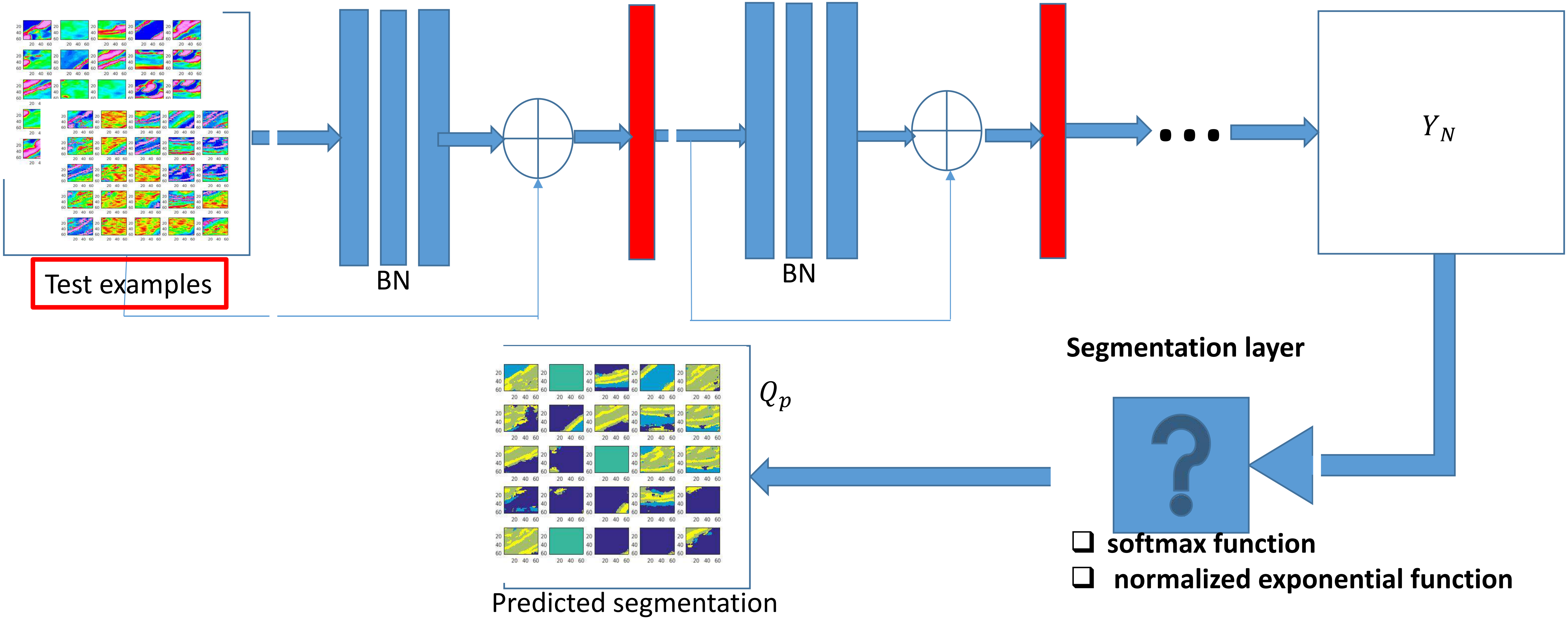
Training



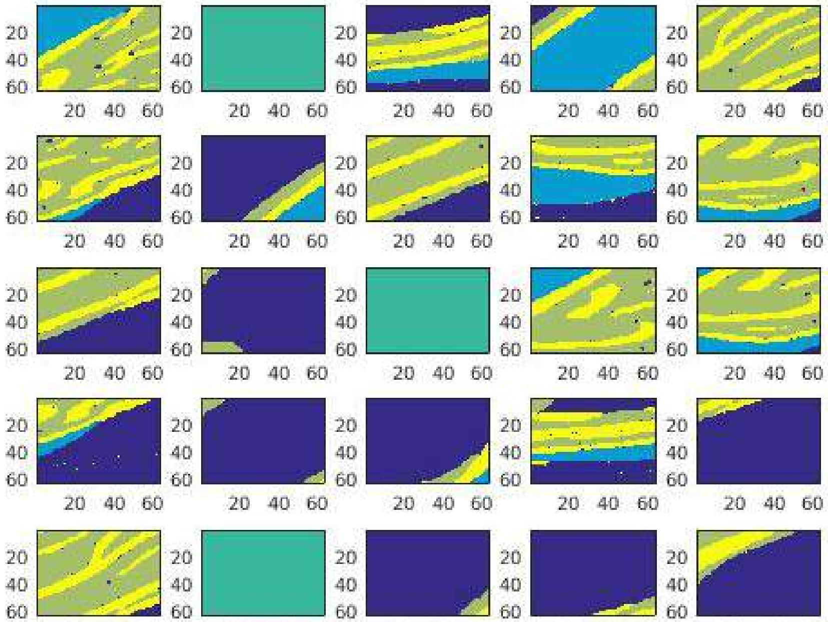
$training\ examples$
Y_N
Segmentation layer
Q
Tune parameters
To be as close as manual segmentation
Cross entropy
for each pixel
Q_p
true segmentation
Predicted segmentation
softmax function
normalized exponential function


A building block
training
conv
relu
inverse
conv
relu
inverse
The output of the forward propagation
Y_N
BN
BN
training examples
Segmentation layer
Q
Tune parameters
To be as close as manual segmentation
Cross entropy for each pixel
Q_p
softmax function
normalized exponential function
true segmentation
Predicted segmentation

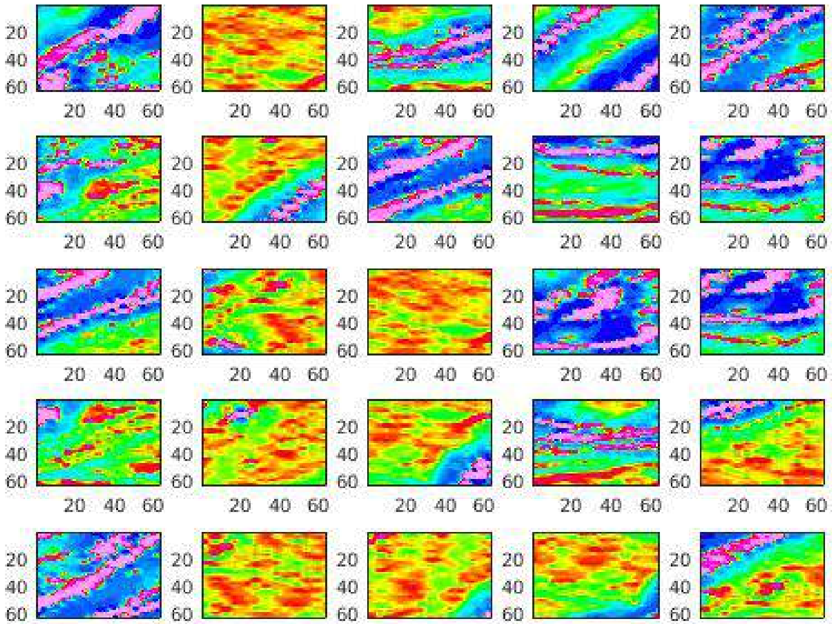
Predicting



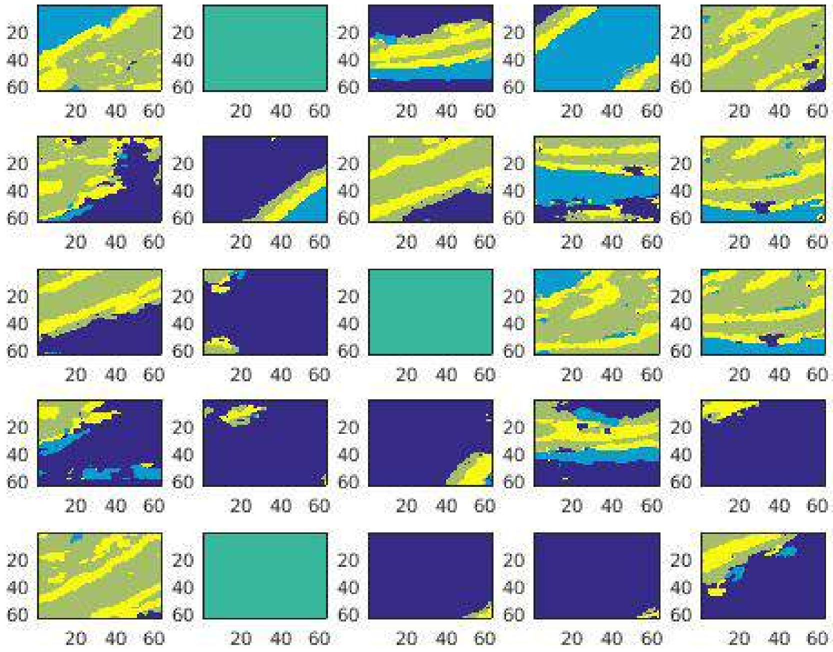
Test examples
BN
BN
Y_N
Segmentation layer
Q_p
Predicted segmentation
softmax function
normalized exponential function



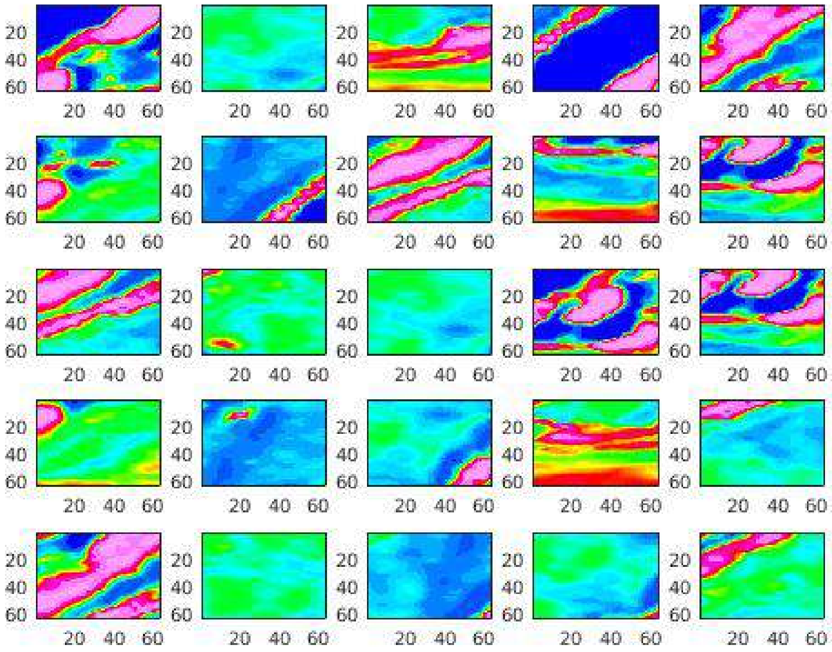
manual segmentation



Test patches: 1VD



automatically segmentation

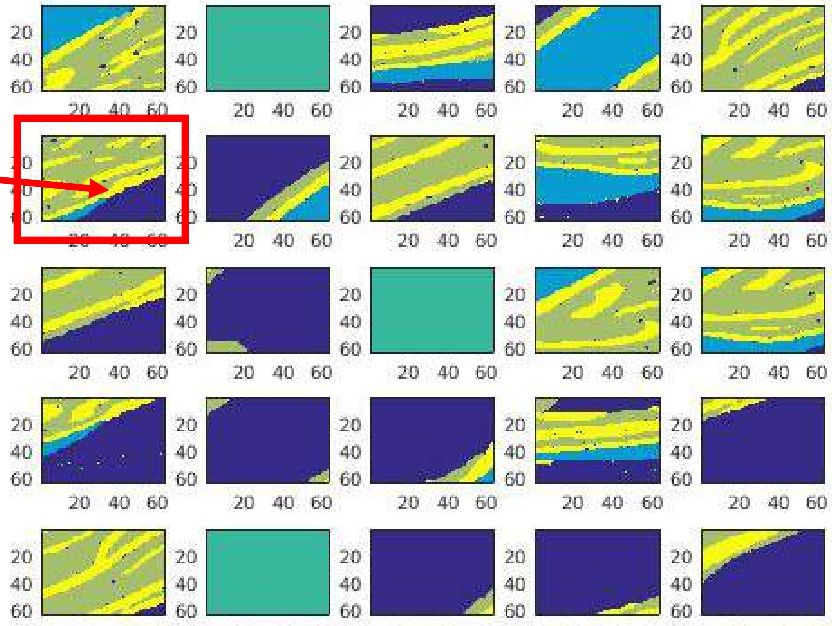


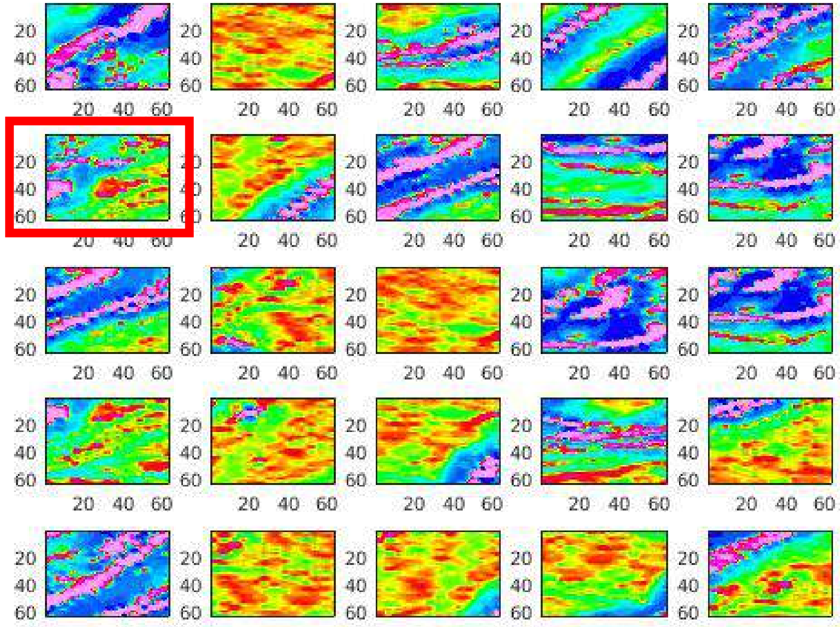
Test patches: TMI

1. Biased

manual segmentation

Long time

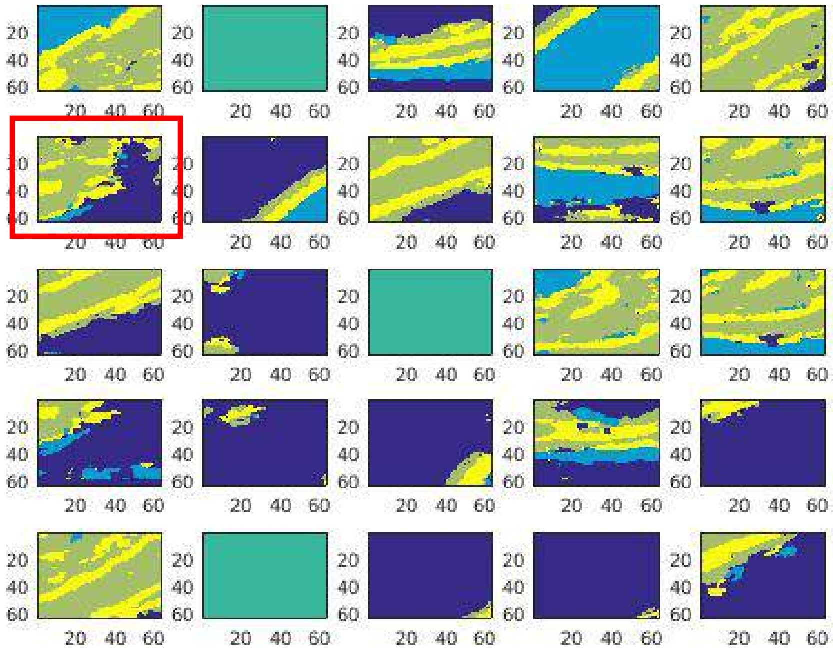


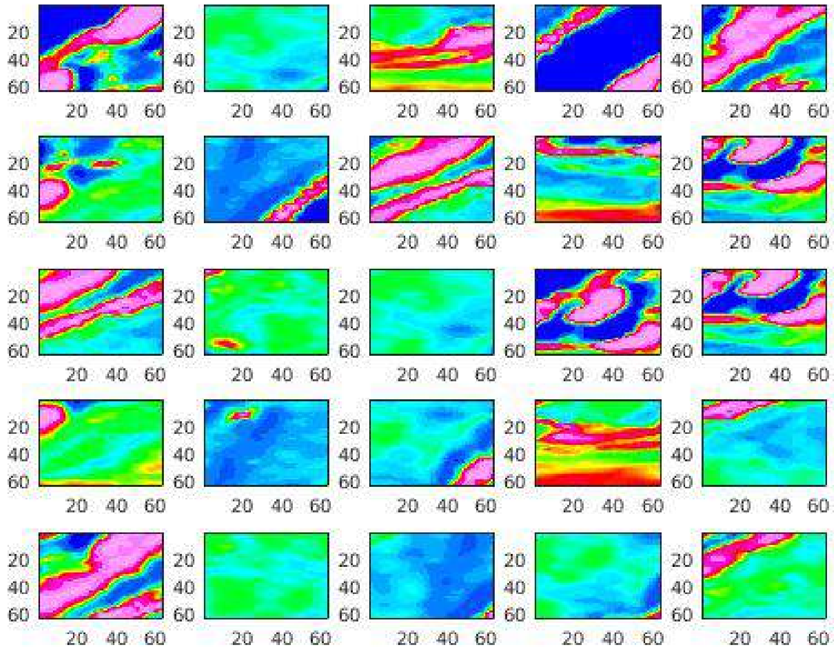


2. Noisy results

automatically segmentation

7 seconds for 200 examples





Summary

- propose a solution to the problem of insufficient annotated magnetic data based on MPS
    - generated realizations are geologically realistic and share the similar spatial continuity with the original data
- propose a new neural network architecture called iCNN to automatically segment magnetic data
- our method is capable of superseding human segmentation in some aspects due to computational efficiency and its ability to identify the geological features