# Image-to-Image Wildfire Detection via Quantum-compatible Variational Segmentation for Remotely-sensed Data

## Earth Science Technology Forum 2023

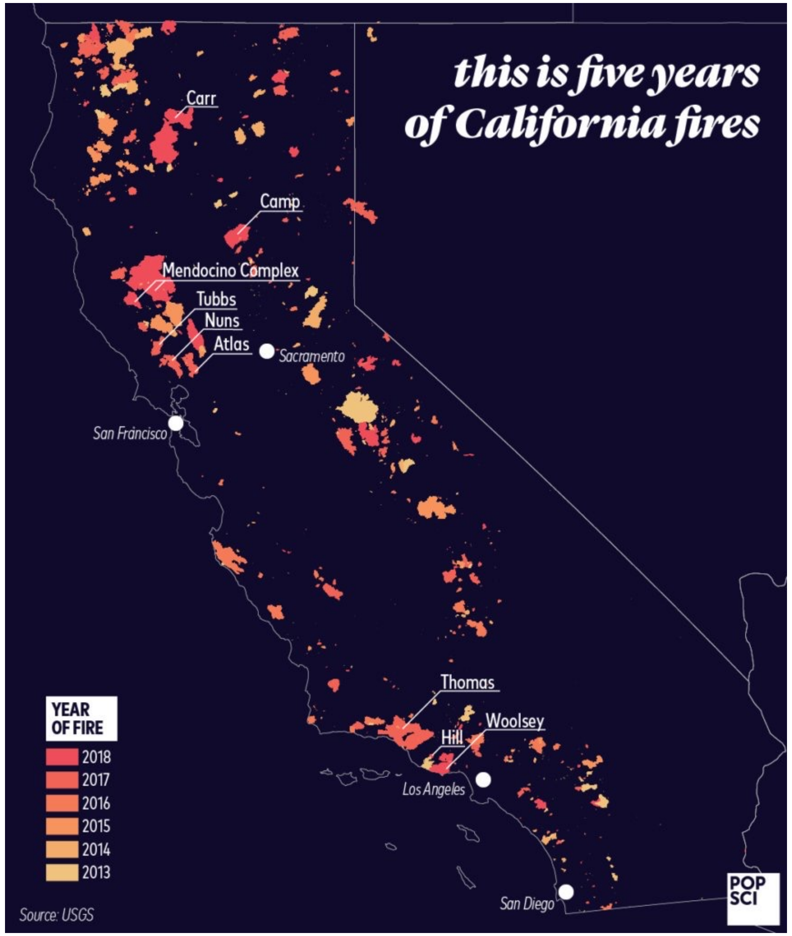
Caltech, Pasadena, CA


Presented by: Ata Akbari Asanjan[2]
PI: Eleanor Rieffel[1]
Co-I: Shon Grabbe[1]
Collaborators: Lucas Brady[1], Nishchay Suri[1], Zoe Gonzalez Izquierdo[1] , P. Aaron Lott[1], Milad Memarzadeh[2]

(1): Quantum Artificial Intelligence Laboratory (QuAIL), NASA Ames Research Center
(2): Data Science Group (DSG), NASA Ames Research Center

# Importance of Wildfire Observation

- Wildfire occurrences have been increasing for the past decades, leaving devastating traces across the globe.
  - *Example: 2018 wildfires in California: $148.5 Bn[1]*
- Proper resource management is crucial in the fight against wildfires.
- Accurate detection is the first step in proper wildfire management.
- Proper machine learning techniques can help discover remote sensing-based information that can help us better characterize wildfires.




Credit: USGS

[1] Wang et al., "Economic footprint of California wildfires in 2018," Nature Sustainability, 4, 252-260 (2021)

# Wildfires are stochastic in nature!

- Like many other natural processes, wildfires are stochastic.
- Wildfire simulations are classified in two categories:
  - **Deterministic**: Assuming wildfire processes are fully resolved.
    - Provides the same outcome every time the model is run for a single wildfire event.
    - Does not account for variability in observations.
  - **Stochastic**: Incorporates the variability of observation.
    - Provides different scenarios every time the model is run for a single wildfire event.
    - Provides a comprehensive statistical understanding for the variability over $N$ runs.
- Thus, deterministic approaches are not optimal for stochastic processes (e.g. wildfire).

Credit: Kevin Maddrey

# Uncertainty in Wildfire Observations

- Uncertainty analysis enables the assessment of reliability and confidence in research results.
- Uncertainty analysis aids in decision-making processes related to resource management, policy development, and risk assessment.
- It helps quantify and communicate the uncertainties associated with observations, measurements, and predictions in Earth science.

- However, uncertainty analysis is not cheap (requires extensive computational and design resources).
- Most uncertainty analysis methods are not designed to run "what-if" scenarios in a low-cost and comprehensive manner.

# Dataset

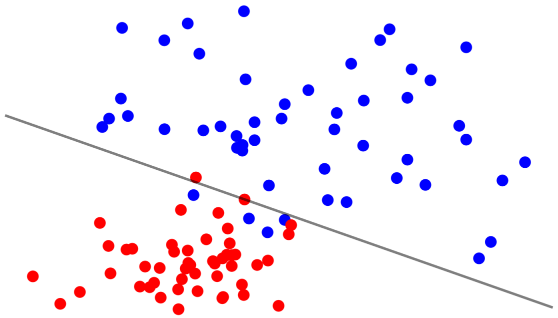
- We used the observations of NASA's Terra and Aqua MODIS for
  - Land/Cloud/Aerosols Boundaries
  - Land/Cloud/Aerosols Properties

- We collected the wildfire mask data from NASA's Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi National Polar-Orbiting Partnership (Suomi NPP).
- We collected 10,000 wildfire samples (with overlapping incidents) over CONUS for the time range of 2018-2020.
- Normalized Difference Vegitation Index (NDVI) is also calculated and included as proxy of vegetation health.
- We included a deviation from mean NDVI accounting for sudden shifts in NDVI in a region.

# Discriminative vs. Generative

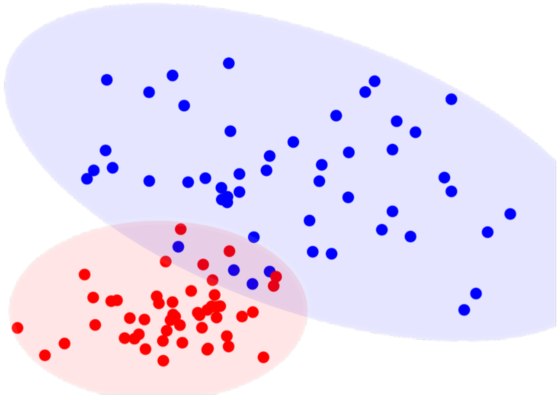
<u>Discriminative modeling:</u>

- In discriminative modeling, we aim to learn a model that discriminat (i.e. predicts) given the inputs. (In probability terms: $p(y \mid X)$)



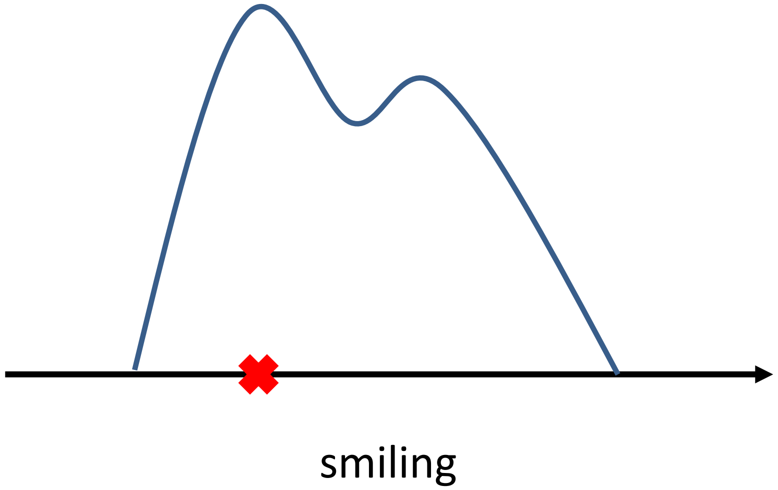
<u>Generative modeling:</u>

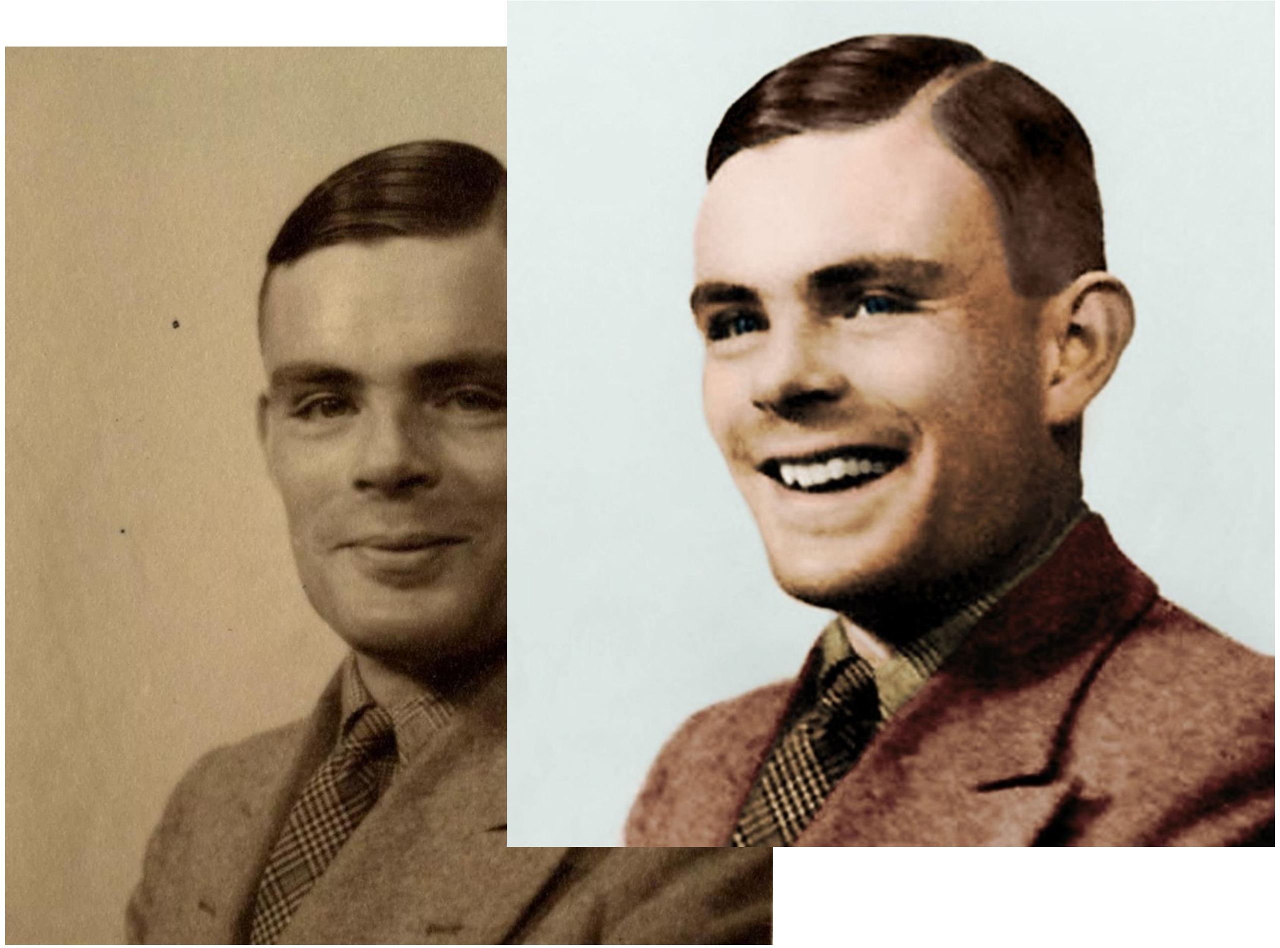
- Generative modeling aims to solve a more general problem. It aims to learn joint distribution over all variables.
(In probability terms: $p(y, X)$ or $p(y \mid X)\, p(X)$)
- A generative model simulates how the data is generated in the real world.



# Generative Modeling based on Statistical Inference

**Statistical Inference** is a learning scheme in which we learn about an unobserved state based on our observations.



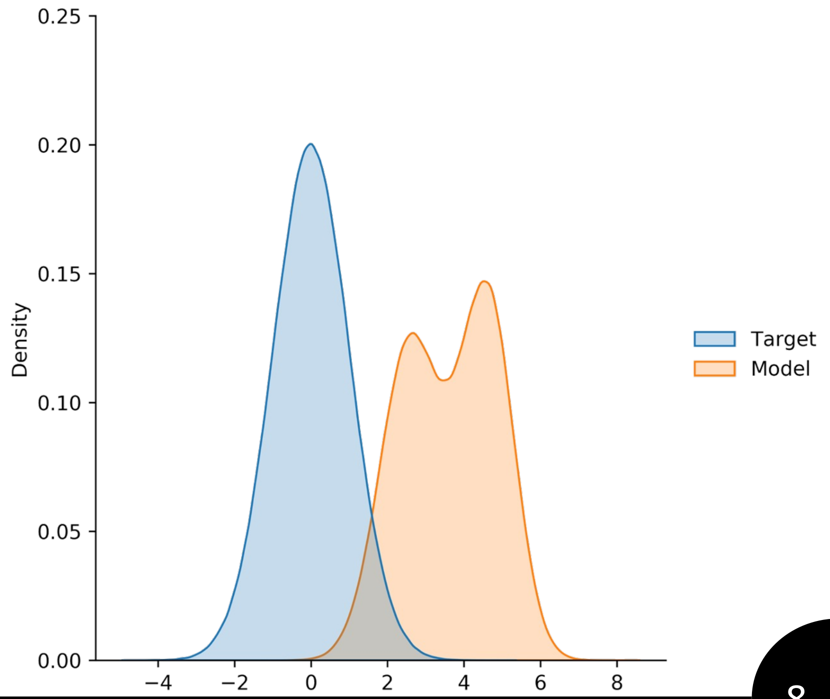
# Variational Inference

Variational Inference suggests that instead of going through all the samples, we assume a distribution (e.g. Gaussian) from distribution family and instead of finding the entire distribution (hard), find the distribution parameters (easier).

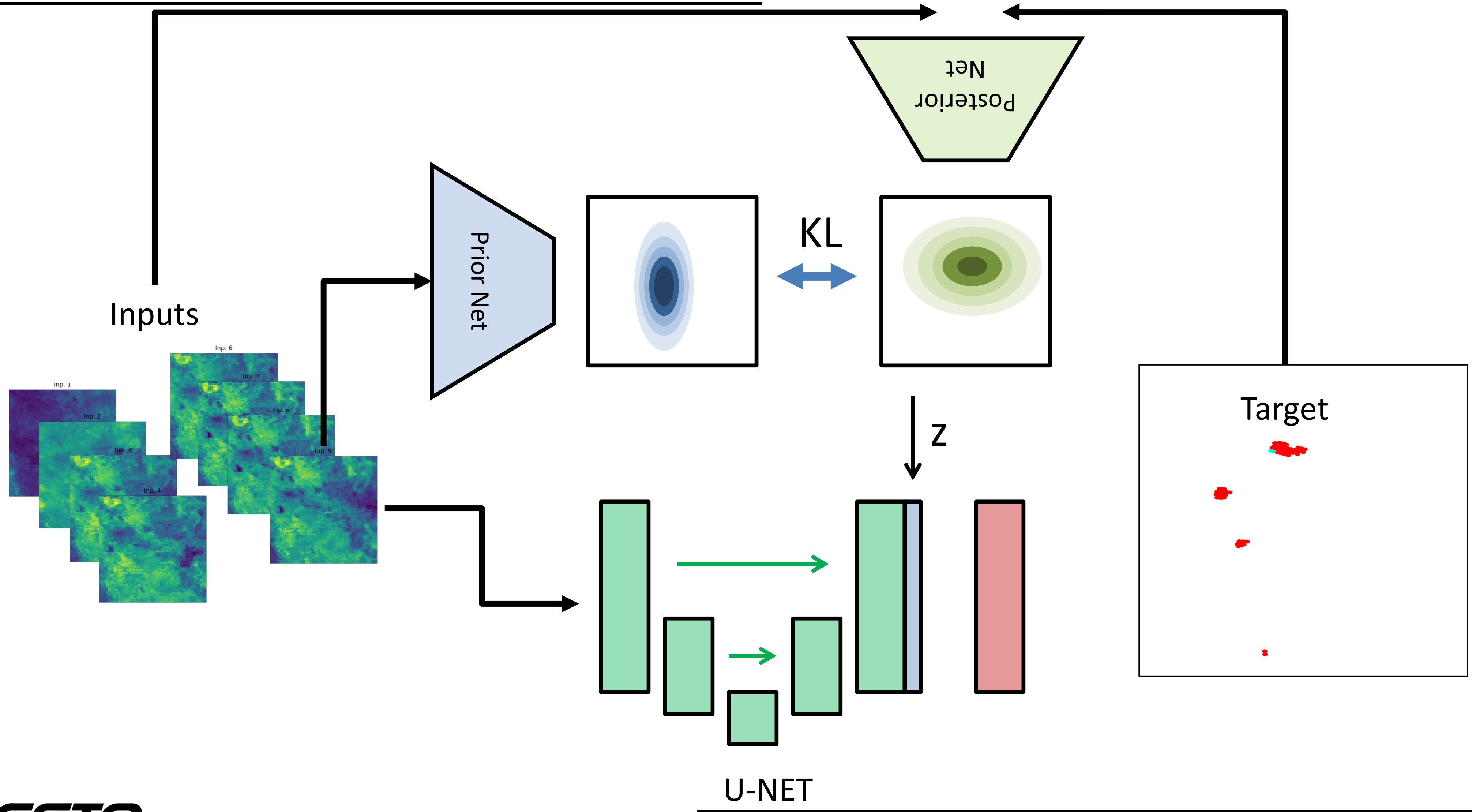
$$p(x) = \int p(x \mid z)\, p(z)\, dz$$

How to measure the closeness of distributions?

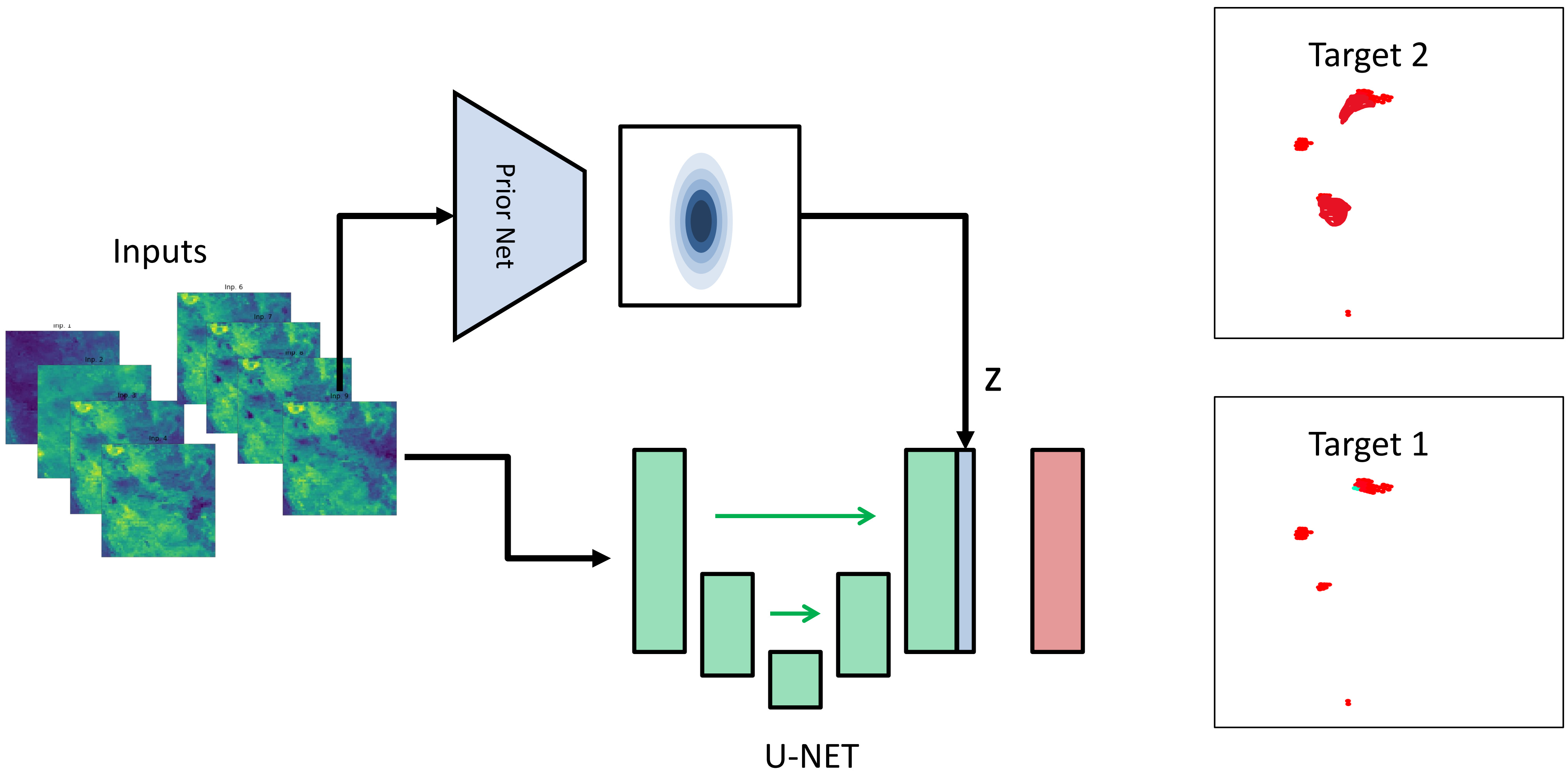
We use a metric called Kullback-Leibler Divergence.



# Probabilistic U-Net – Training Mode



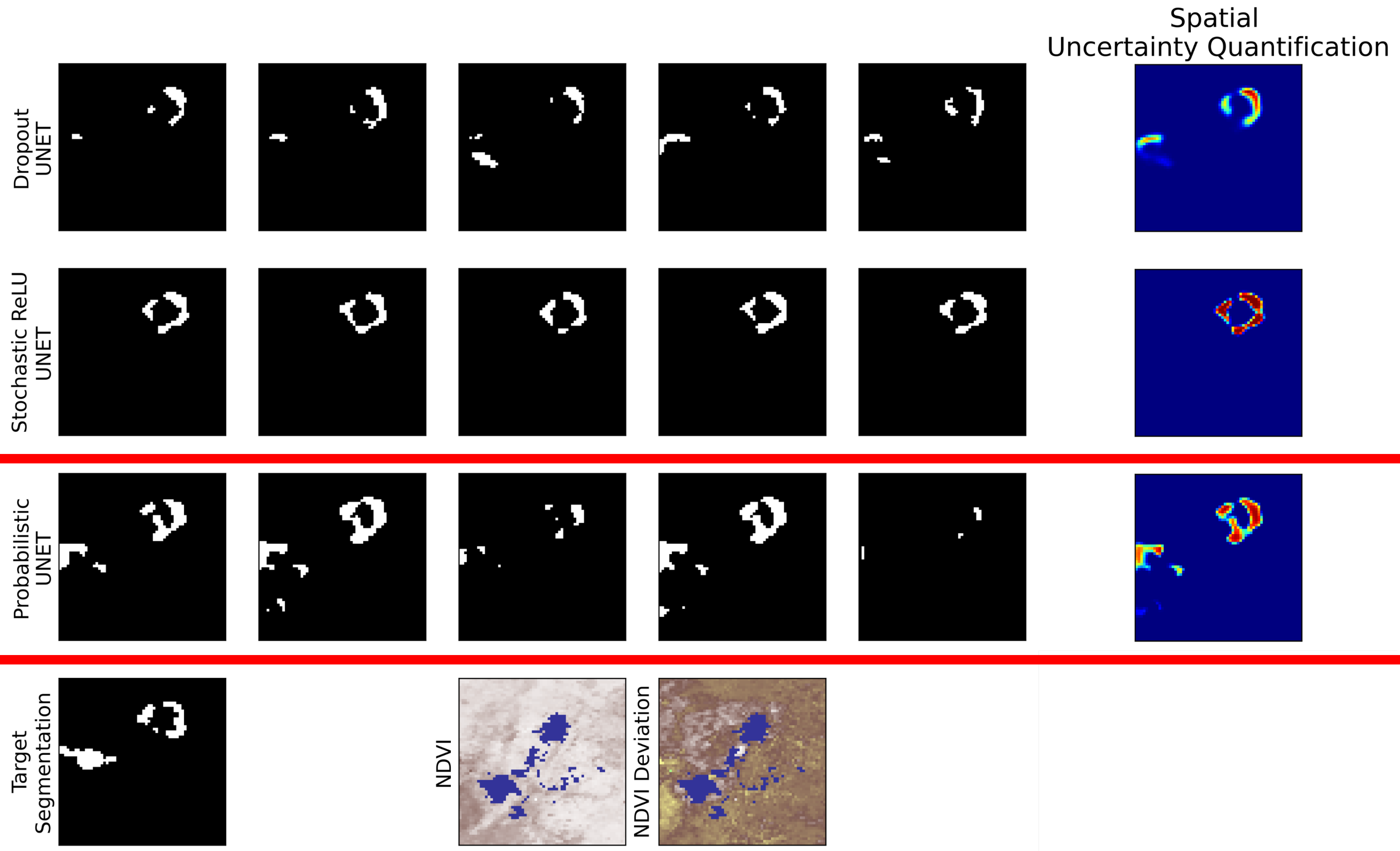
# Probabilistic U-Net – Inference Mode



# Quantum Advantage

- Probabilistic U-NET's performance depend on quality of latent space samples.
- We improved it by relaxing the variation inference assumption (i.e. latent space is a Multivariate Gaussian distribution).
- In order to relax the posterior assumption, we can replace the posterior latent space with an iterative process such as Restricted Boltzmann Machine (RBM).
- The RBM allows parallel Gibbs sampling which results in more accurate prior characterization.
- This way we are joining the best of both worlds (Variational Inference & MCMC) to generate more accurate latent samples and thus, more realistic scenarios for wildfire detection.

# Visual Comparison



# Statistical Comparison

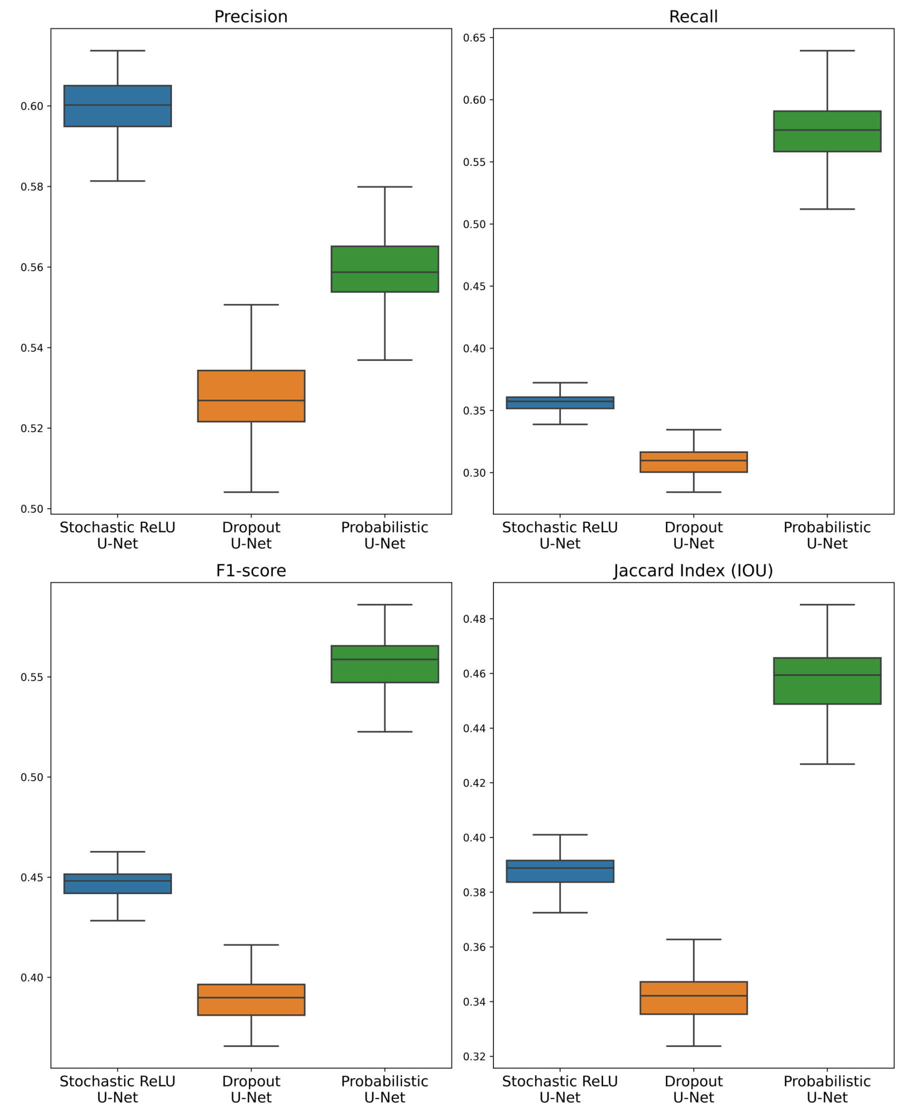
$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 score} = \frac{2 \times Precision \times Recall}{Precision \times Recall}$$

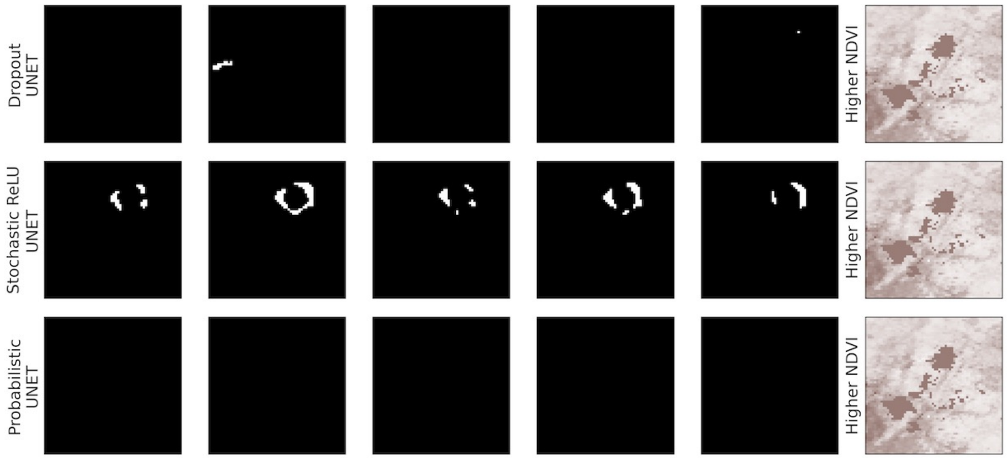
$$\text{Jaccard score} = \frac{A \cap B}{A \cup B}$$

| | | Truth | |
|---|---|---|---|
| | | Fire | No Fire |
| Prediction | Fire | TP | FP |
| | No Fire | FN | TN |

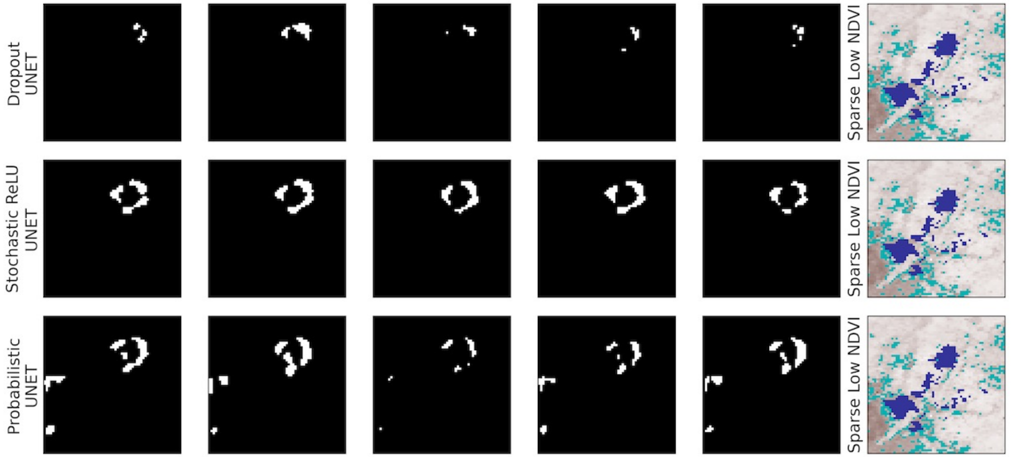


# What-if Scenarios

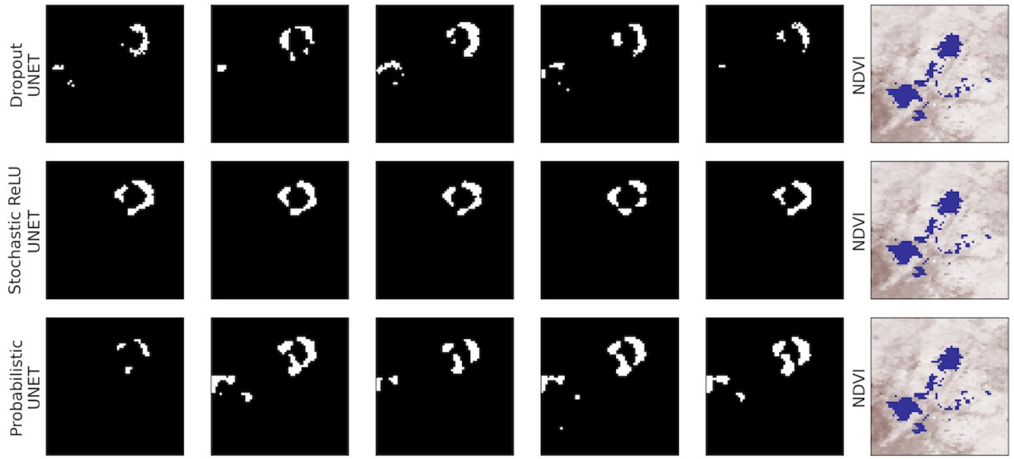
## Healthy Vegetation




## Sparse unhealthy Vegetation




## Normal Vegetation




## Very unhealthy/No Vegetation

# Highlights

- Generative machine learning can improve our understanding of wildfire processes and offer a promising approach for wildfire detection and uncertainty quantification.
- The proposed approach demonstrates the ability to generate stochastic wildfire detections, enabling comprehensive uncertainty quantification for individual and collective events.
- Incorporating uncertainty analysis in wildfire detection enhances decision-making capabilities for authorities, aiding in effective mitigation and prevention strategies.
- The findings highlight the potential of generative machine learning in advancing wildfire detection and decision support systems, contributing to improved wildfire management and public safety.

# Acknowledgement & References

- This work was supported by the NASA ESTO Advanced Information Systems Technology Program through grant AIST-QRS-21.

- Akbari Asanjan, A., Memarzadeh, M., Lott, P. A., Rieffel, E., & Grabbe, S. (2023). Probabilistic Wildfire Segmentation Using Supervised Deep Generative Model from Satellite Imagery. Remote Sensing, 15(11), 2718.
- Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J. R., Maier-Hein, K., ... & Ronneberger, O. (2018). A probabilistic u-net for segmentation of ambiguous images. *Advances in neural information processing systems*, *31*.
- Winci, W., Buffoni, L., Sadeghi, H., Khoshaman, A., Andriyash, E., & Amin, M. H. (2020). A path towards quantum advantage in training deep generative models with quantum annealers. *Machine Learning: Science and Technology*, *1*(4), 045028.
- Khoshaman, A., Vinci, W., Denis, B., Andriyash, E., Sadeghi, H., & Amin, M. H. (2018). Quantum variational autoencoder. *Quantum Science and Technology*, *4*(1), 014001.

Thank you very much for your attention!

Questions?

Ata Akbari Asanjan

*ata.akbariasanjan@nasa.gov*

# Generative Modeling based on Probabilistic Inference

<u>Bayes rule:</u>

$$p(z \mid x) = \frac{p(x \mid z)\, p(z)}{p(x)} = \frac{p(x, z)}{p(x)}$$

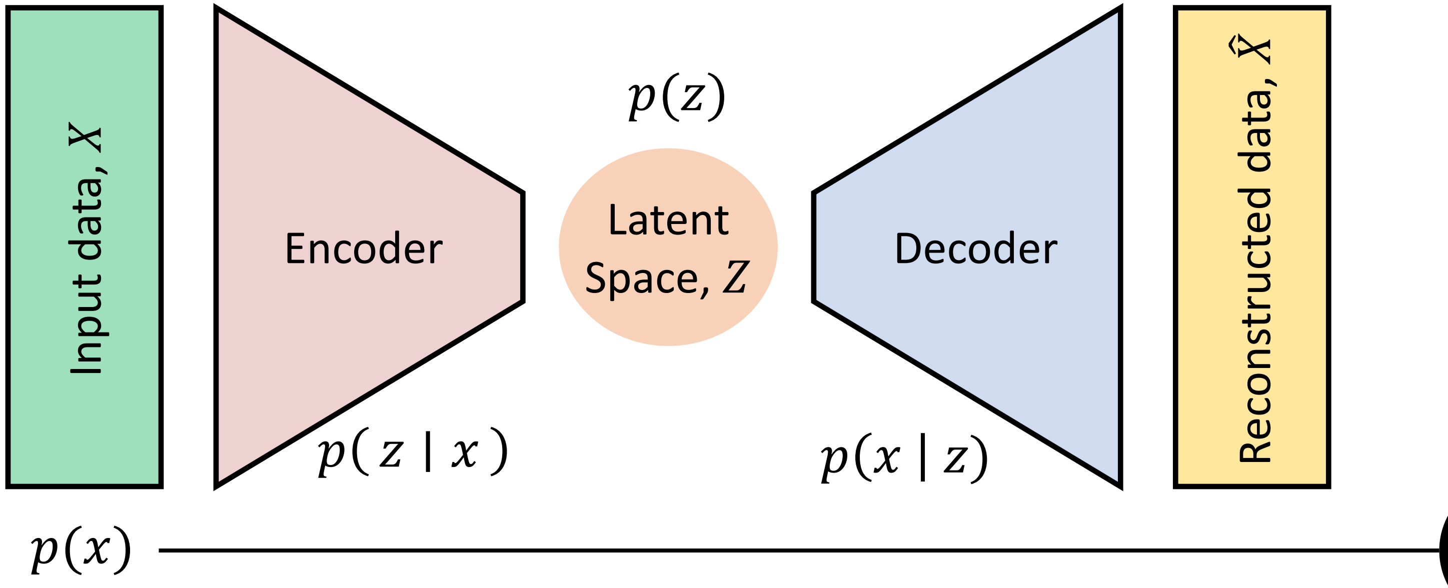
- $p(x)$ is data distribution or Evidence.
  - (In discriminative models, we rather focus on conditional probability $p(y|x)$ and neglect the unconditional probability $p(x)$.
- $p(z)$ is the prior distribution.
- $p(x \mid z)$ is the likelihood.
- $p(z \mid x)$ is posterior distribution.

# Probabilistic Inference – Unsupervised Form

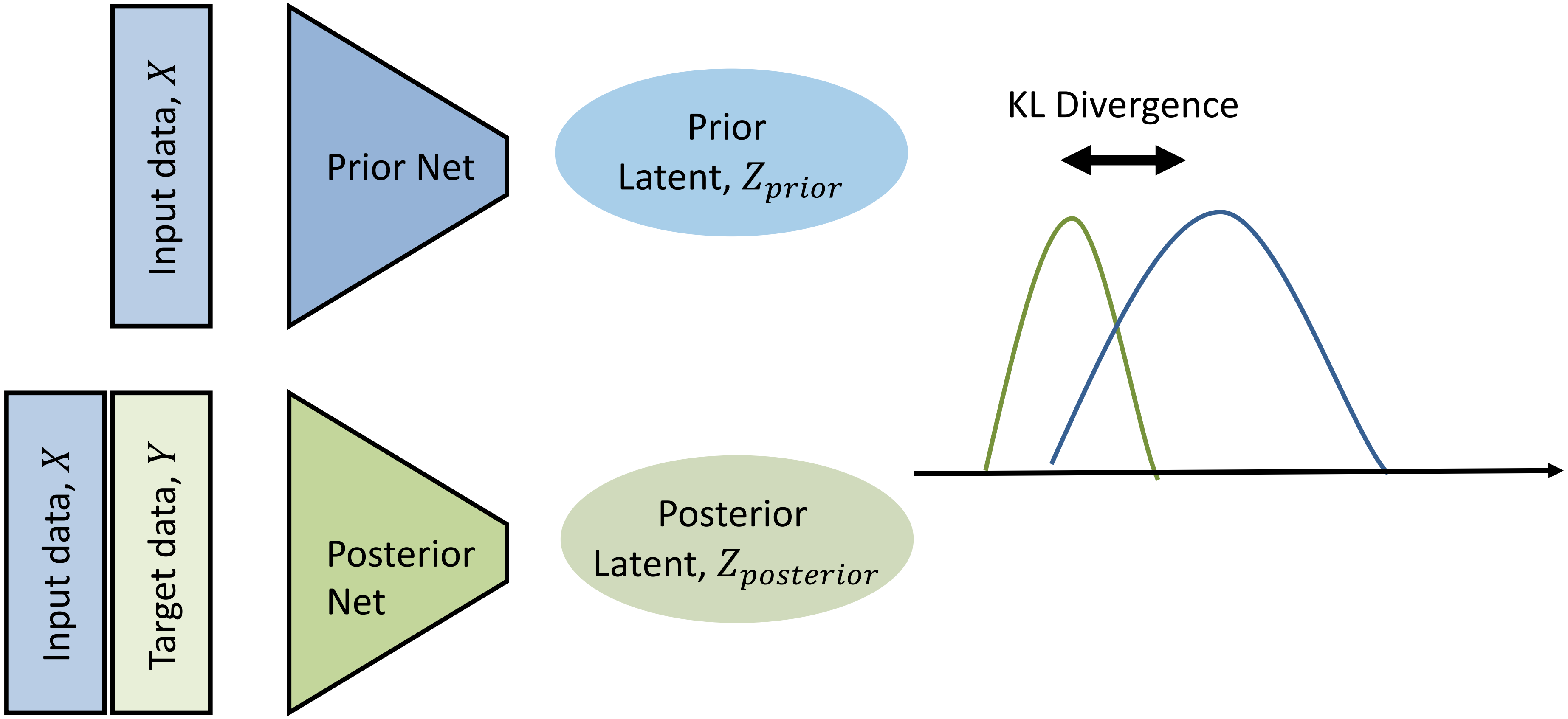
Bayes rule:

$$p(z \mid x) = \frac{p(x \mid z)\, p(z)}{p(x)} = \frac{p(x, z)}{p(x)}$$

In unsupervised variational inference we assume a family of distributions for the prior and force the model to learn the best distribution parameters that match the data.



# Probabilistic Inference – Supervised Form



# Probabilistic Inference – Supervised Form

- Probabilistic U-Net is a great approach for capturing variations in a supervised fashion.
- However, it can be further improved by relaxing the variation inference assumption (i.e. latent space is a Multivariate Gaussian distribution).
- In order to relax the prior assumption, we can replace the prior latent space with an iterative process such as Restricted Boltzmann Machine (RBM).
- The RBM allows parallel Gibbs sampling which results in more accurate prior characterization.
- This way we are joining the best of both worlds (Variational Inference & MCMC) to generate more accurate latent samples and thus, more realistic scenarios for wildfire detection.

# Statistical Inference

Bayes rule:

$$p(z \mid x) = \frac{p(x \mid z)\, p(z)}{p(x)}$$

- Solving the Bayesian inference in the previous slide is often hard close to not possible.
- This becomes worst with larger dimensionality in data (e.g. Image, time series).

$$p(x) = \int p(x \mid z)\, p(z)\, dz$$

Solutions:

1. Variational Inference: Moderate accuracy, Fast
2. Markov Chain Monte Carlo: Good accuracy, Very slow

# Monte Carlo Markov Chain

- MCMC is a generic method of sampling from a high-dimensional probability distribution.
- By sampling, we gain better knowledge of the entire probability distribution landscape.
- *As we sample more from a distribution,*

*we learn more about the distribution!*

- MCMC includes many variations
  - **Metropolis-Hasting**: Uses proposal density

  & acceptance/rejection method for new samples.
  - **Gibbs**: Uses conditional distributions for new samples. (Good for complex high-dimensional target distributions)

# Gibbs Sampling

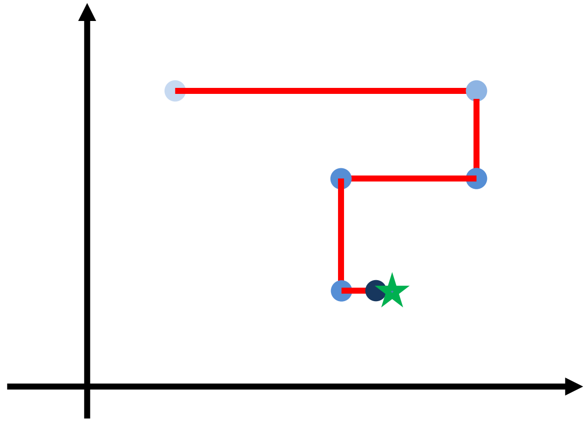
- Gibbs sampling breaks down the sampling process of a complex high-dimensional target distribution, into simpler, easy-to-sample conditional distributions.
  - Example: Imagine we have a $N$-d target distribution

$$P(x_1, x_2, x_3, \dots, x_N)$$

  - Drawing samples from this distribution is hard if we don't have the joint probability function.
  - Instead, we freeze all but one dimension and calculate a conditional probability. e.g.;

$$P(x_1 \mid x_2, x_3, \dots, x_N)$$

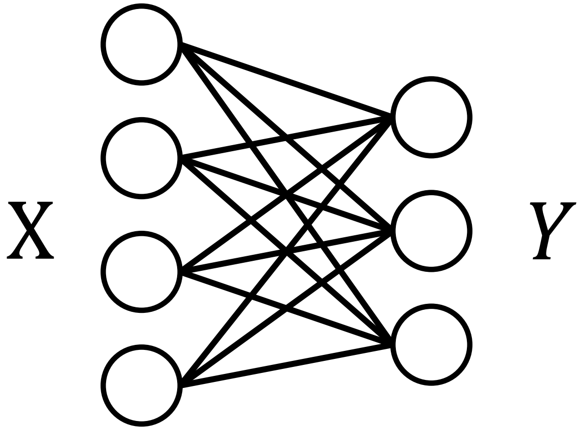
  - Then we start from a random location, update each dimension based on other given dimensions and conditional probability



# Gibbs Sampling in the form of ML

- Gibbs sampling can be implemented as a machine learning model.
- Imagine we have two variables $X$ and $Y$.
- In order to sample from the joint $\mathrm{P}(X, Y)$ distribution, all we need is to have $\mathrm{P}(X \mid Y)$ and $\mathrm{P}(\mathrm{Y} \mid X)$.
- We can define a model that gives the conditional distributions: **Restricted Boltzmann Machine (RBM)**!
- RBM learns conditional distributions via negative log-likelihood.
- Gibbs sampler uses conditional distributions to refine samples.
- This mechanism learns a Boltzmann distribution of $X$ and $Y$.

$$P(X,Y) = \frac{e^{-E(x)}}{\sum_{X,Y} e^{-E(x,y)}}$$



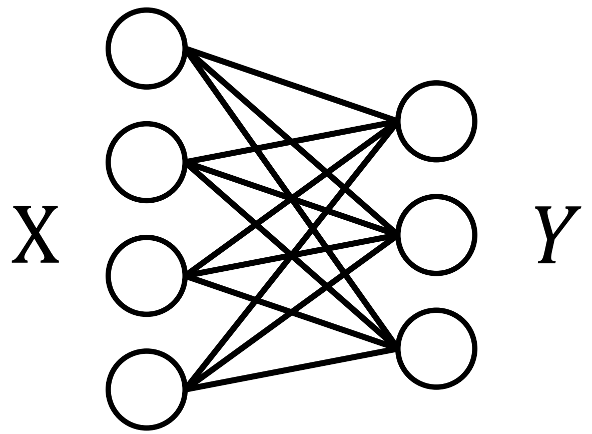
# RBM: An energy-based model

- This mechanism learns a Boltzmann distribution of $X$.

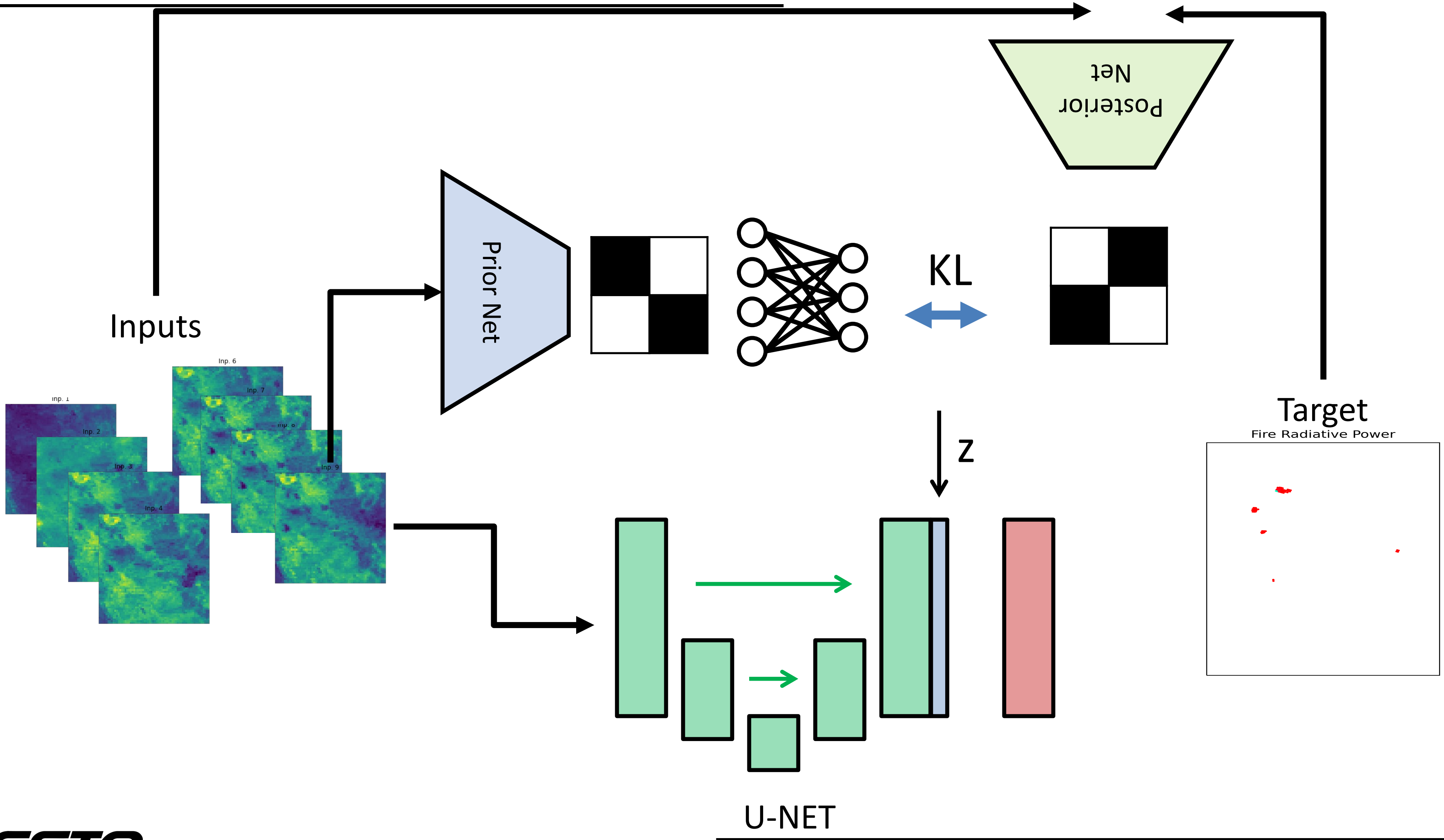
$$P(X) = \sum_{Y} \frac{e^{-E(x,y)}}{\sum_{X,Y} e^{-E(x,y)}}$$

- Energy term $E(x, y)$ can be represented by

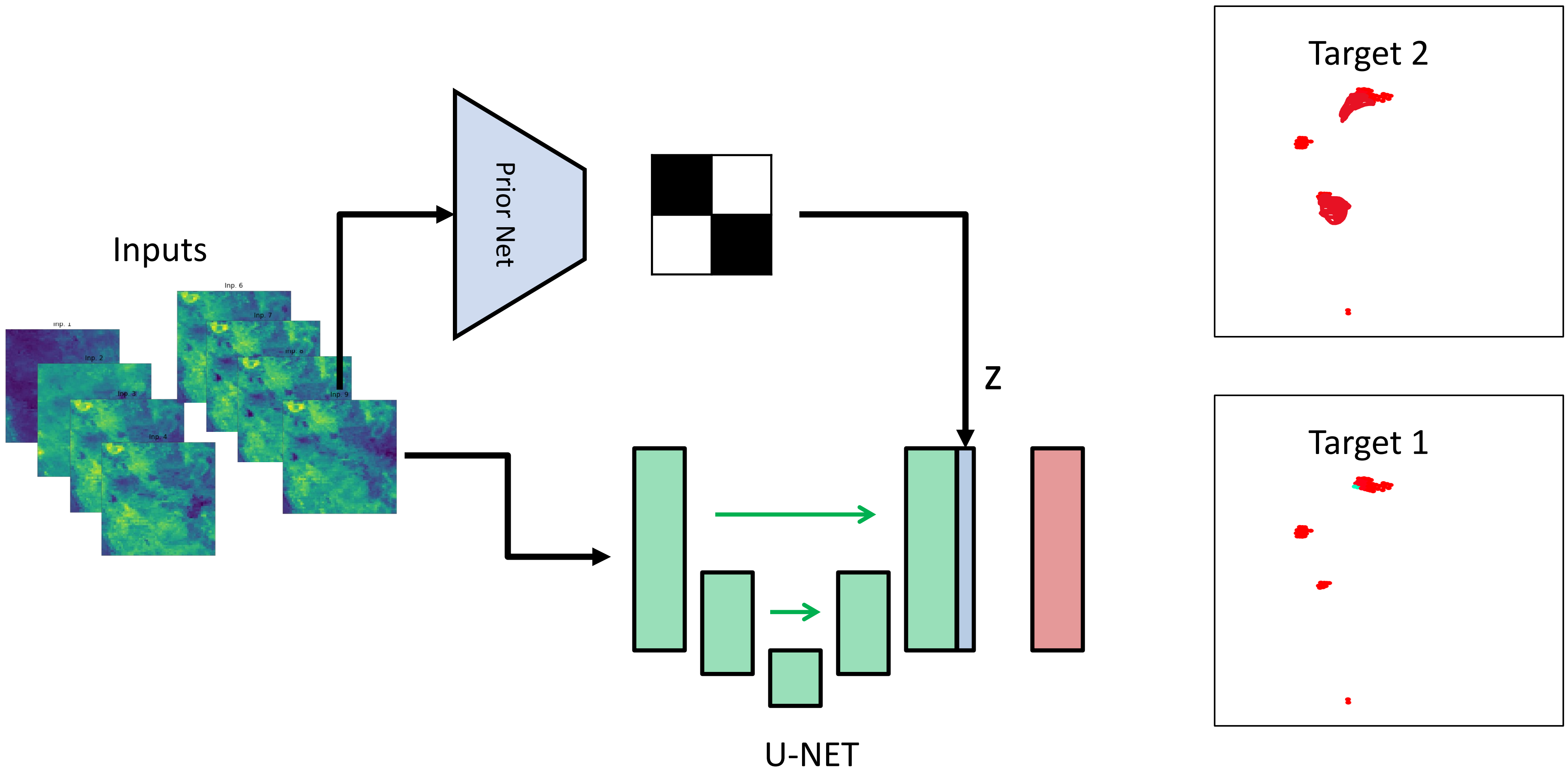
$$E(x, y) = -\sum_{i \in X} x_i \, b_i^X - \sum_{j \in Y} y_j b_j^Y - \sum_{i \in X} \sum_{j \in Y} x_i y_j w_{ij}$$



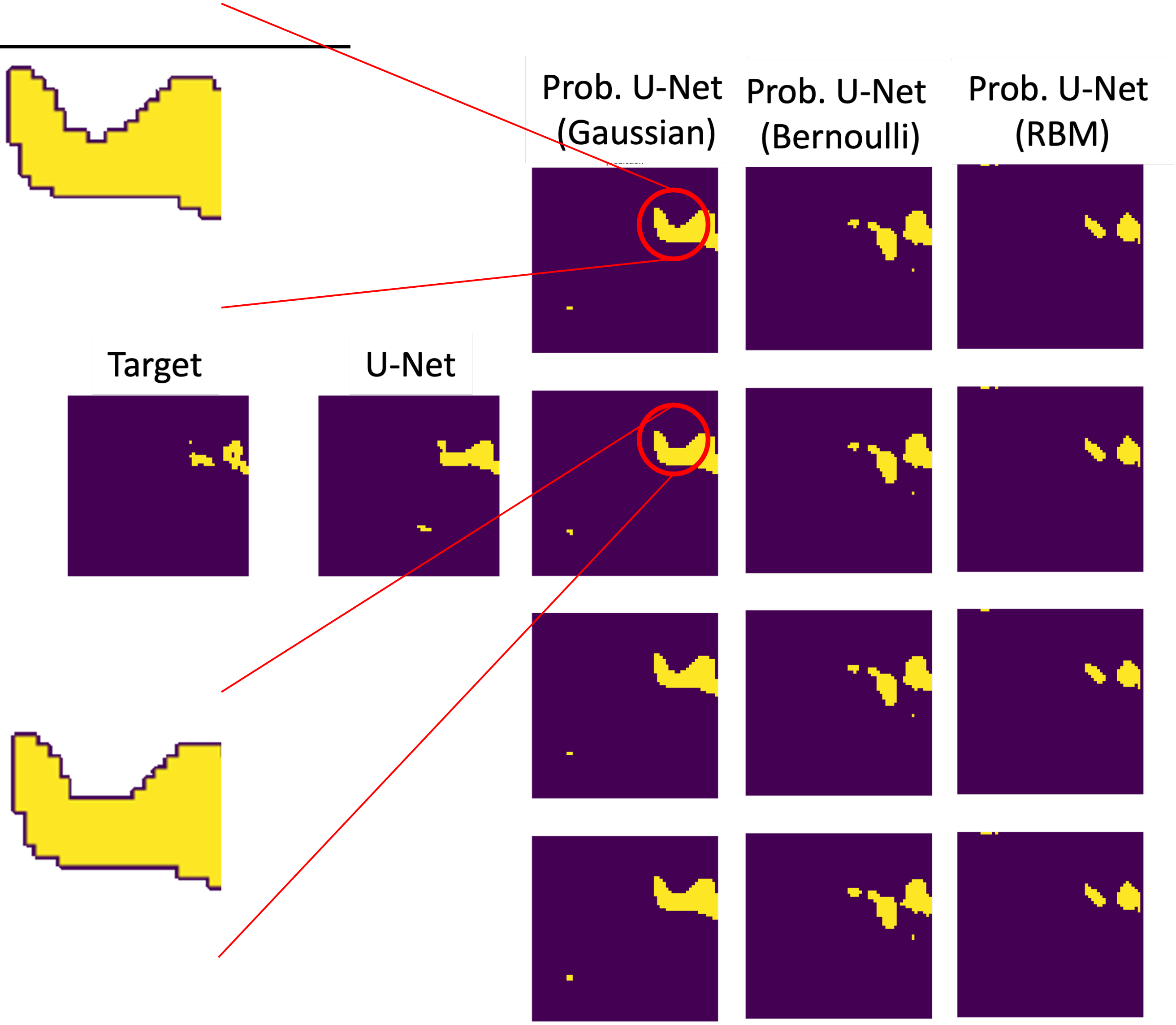
# Combining RBM with Probabilistic U-Net – Training mode



# Combining RBM with Probabilistic U-Net – Inference Mode



# Visual Results



# Performance Metrics

| | U-Net | Prob. U-Net (Gaussian) | Prob. U-Net (Bernoulli) | Prob. U-Net (RBM) |
|---|---|---|---|---|
| Precision | 0.536 | 0.431 | 0.235 | **0.654** |
| Recall | **0.987** | 0.955 | 0.752 | 0.473 |
| F1 score | **0.695** | 0.594 | 0.358 | 0.549 |
| Jaccard score | **0.532** | 0.422 | 0.318 | 0.378 |

Precision = $\frac{TP}{TP + FP}$

Recall = $\frac{TP}{TP + FN}$

F1 score = $\frac{2 \times Precision \times Recall}{Precision \times Recall}$

Jaccard score = $\frac{A \cap B}{A \cup B}$

| | | Truth | |
|---|---|---|---|
| | | Fire | No Fire |
| Prediction | Fire | TP | FP |
| | No Fire | FN | TN |

# Advantages of the proposed approach

- Probabilistic U-Net with Boltzmann latent space is more generalized than its alike with Gaussian latent.
- Discrete latent space will help the model in efficient learning of latent configurations.
- RBM acts as a connection door between the classical and quantum computation realms.

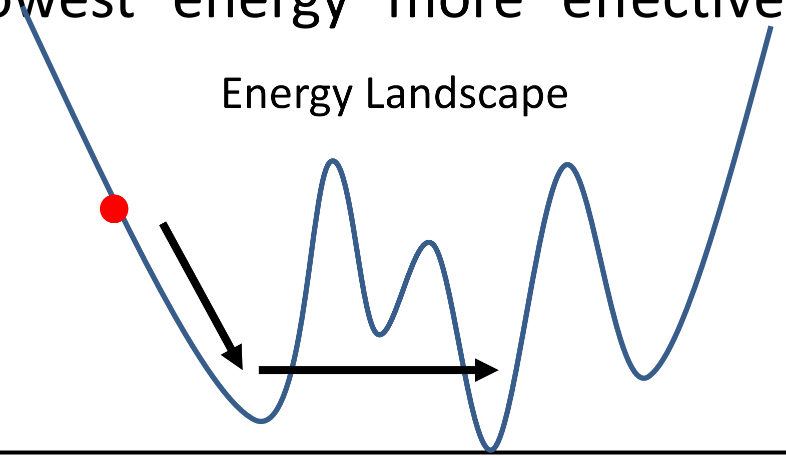
- ***Question**: How RBM connects classical and quantum computations?*
  - RBM uses $e^{-E(x)}$ to define probability, thus;

$$E(x) \propto \frac{1}{p}$$

  - Because of this property, we can look for lower energy to find higher probability samples.

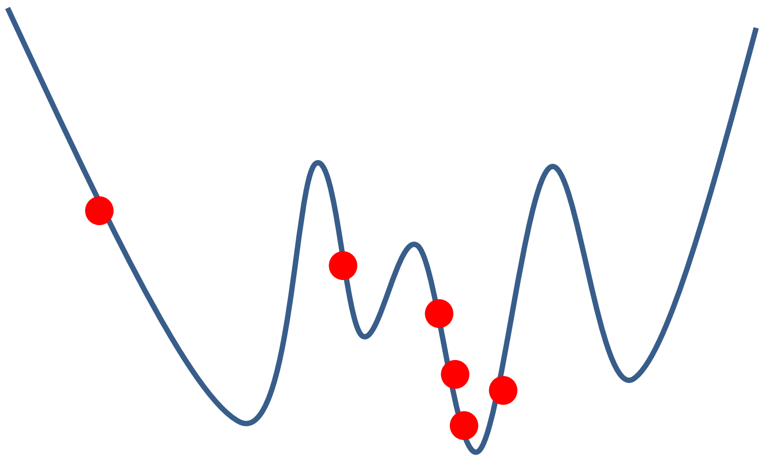
# Quantum Computation

- Quantum computing is a rapidly emerging technology based on quantum mechanics.
- Multiple applications, such as optimization and sampling, have been introduced and are expected to surpass the classical computers' performances.
- Quantum annealing is a proposed optimization method for finding the lowest energy (best answer).
- We start from an initial Hamiltonian state and slowly move toward problem Hamiltonian (solution).
- Theoretically, quantum computer can find the lowest energy more effectively due to tunneling effect.

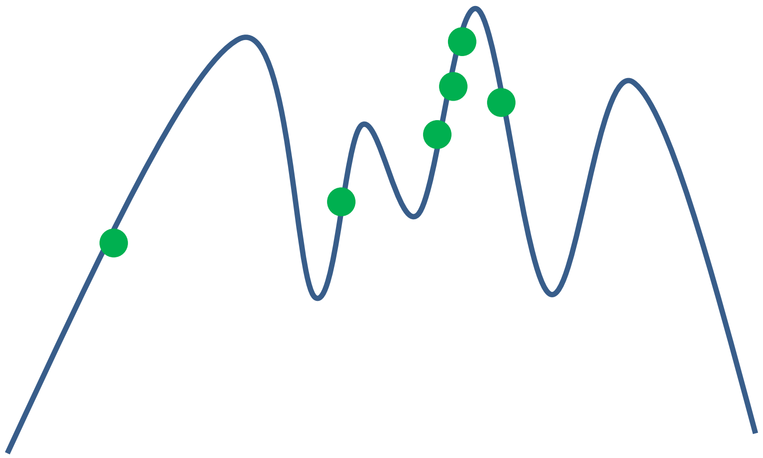


# Bridge between Quantum  and Classical Computation

- We can use this property in sampling to find the best Boltzmann distribution.
- We do MCMC in Energy landscape to find the lowest energy point. That is equivalent to doing MCMC on Boltzmann distribution.
- This approach is expected to perform better because of Quantum computer's effective and fast sampling.
- The results are expected to be more accurate and simultaneous.



Energy Landscape



Probability Distribution