# Learning Gradients of Convex Functions with Monotone Gradient Networks

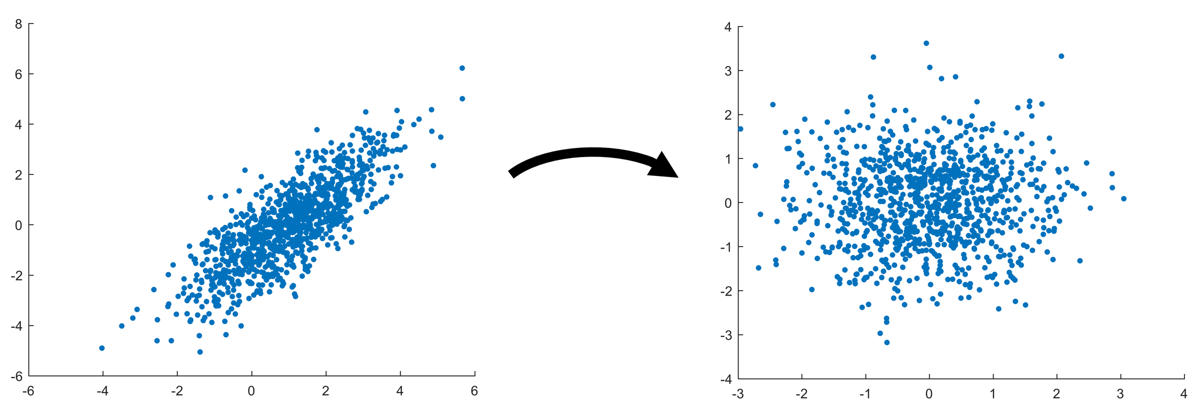
*Shreyas Chaudhari*, Srinivasa Pranav*, José M.F. Moura*

Carnegie Mellon University
Electrical & Computer Engineering

ICASSP 2023

## OVERVIEW

- Convex formulations of signal processing and inverse problems often require domain expertise
- *Recent Trend*: Learn a suitable convex objective that is optimized at test time
- We directly learn gradients of such convex functions
- Diverse applications include optimal transport:



$$\inf_{g: g(x)\sim p_Y} \mathbb{E}_{x\sim p_X} ||x - g(x)||_2^2$$

- *Brenier's Theorem* [1]: unique optimal $g$ is the monotone gradient of a convex function

*Key Contributions*

- We propose *Monotone Gradient Networks* for learning gradients of convex functions
- Our neural networks are simpler to train and achieve better performance than prior approaches [2, 3]

## PROBLEM STATEMENT

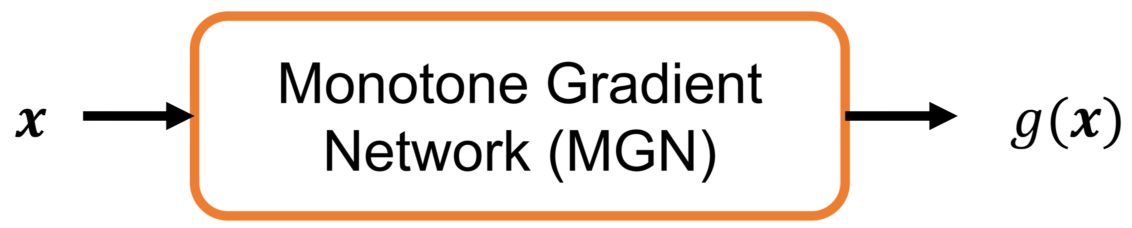
- Goal: Define a neural network $g(\boldsymbol{x})$ that is the gradient of a convex and twice differentiable $f(\boldsymbol{x})$
- Differentiable $f(\boldsymbol{x})$ is convex *iff* its gradient $g(\boldsymbol{x})$ is monotone:

$$\langle g(\boldsymbol{x}) - g(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle \geq 0 \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f)$$

- *Twice* differentiable $f(\boldsymbol{x})$ is convex *iff* its Hessian $\boldsymbol{H}_f(\boldsymbol{x})$ is positive semidefinite (PSD):

$$\boldsymbol{H}_f(\boldsymbol{x}) = \boldsymbol{J}_g(\boldsymbol{x}) \succcurlyeq 0 \quad \forall \boldsymbol{x} \in \mathrm{dom}(f)$$

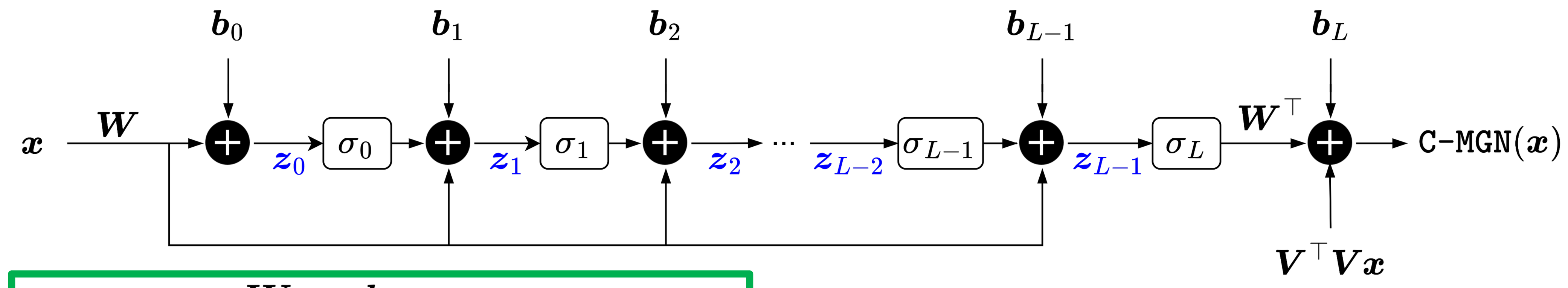
We propose two neural network architectures to parameterize $g(\boldsymbol{x})$




MGN's Jacobian is guaranteed to be PSD

* Equal Contribution
and U.S. National Science Foundation Graduate Research Fellows (NSF GRFP)

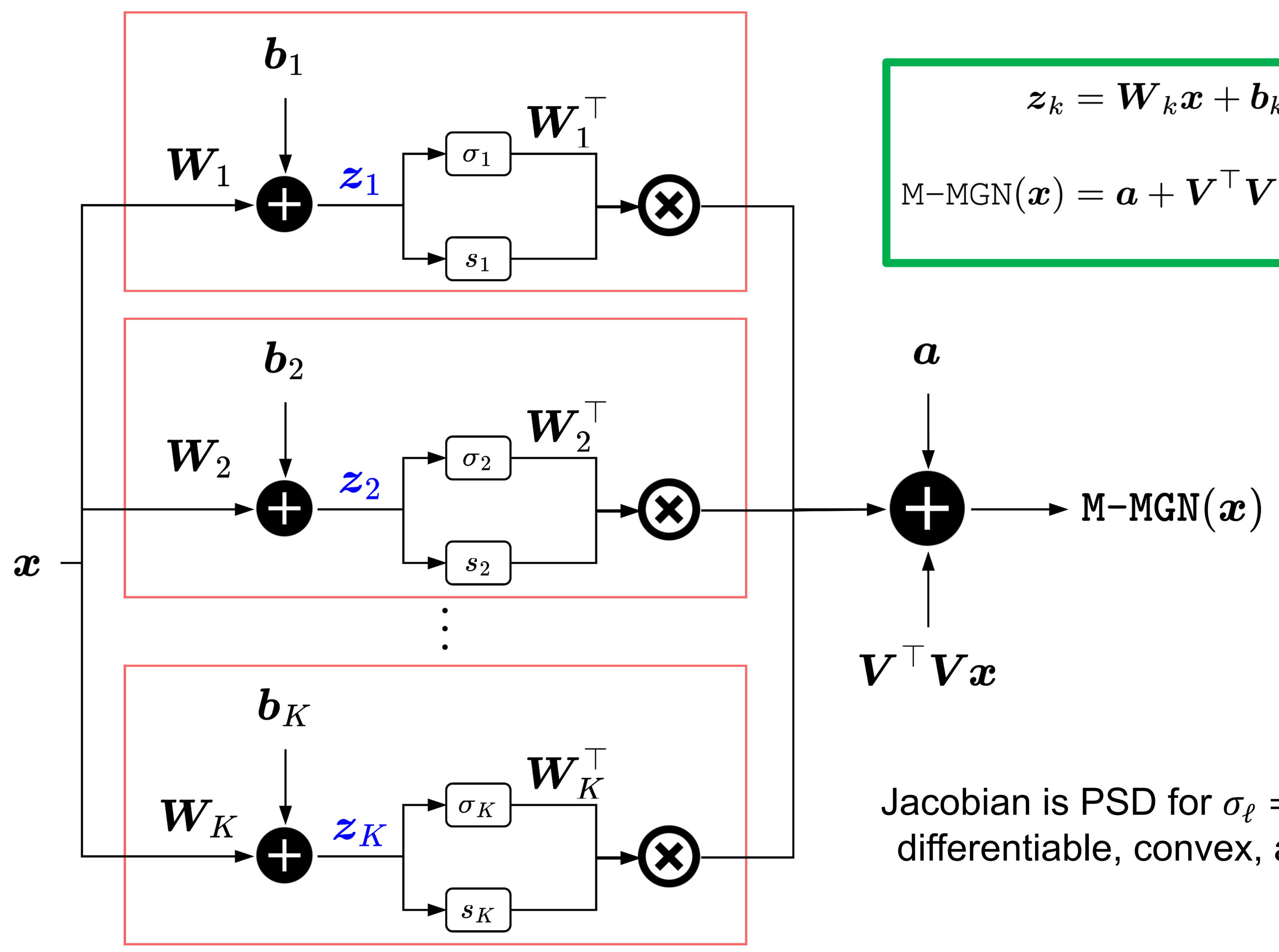
## CASCADED MONOTONE GRADIENT NETWORK (C-MGN)




$$\boldsymbol{z}_0 = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}_0$$
$$\boldsymbol{z}_\ell = \boldsymbol{W}\boldsymbol{x} + \sigma_\ell(\boldsymbol{z}_{\ell-1}) + \boldsymbol{b}_\ell$$
$$\text{C-MGN}(\boldsymbol{x}) = \boldsymbol{W}^\top \sigma_L(\boldsymbol{z}_{L-1}) + \boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{x} + \boldsymbol{b}_L$$

Jacobian is PSD if each $\sigma_\ell(\cdot)$ is an element-wise monotonically increasing function

## MODULAR MONOTONE GRADIENT NETWORK (M-MGN)



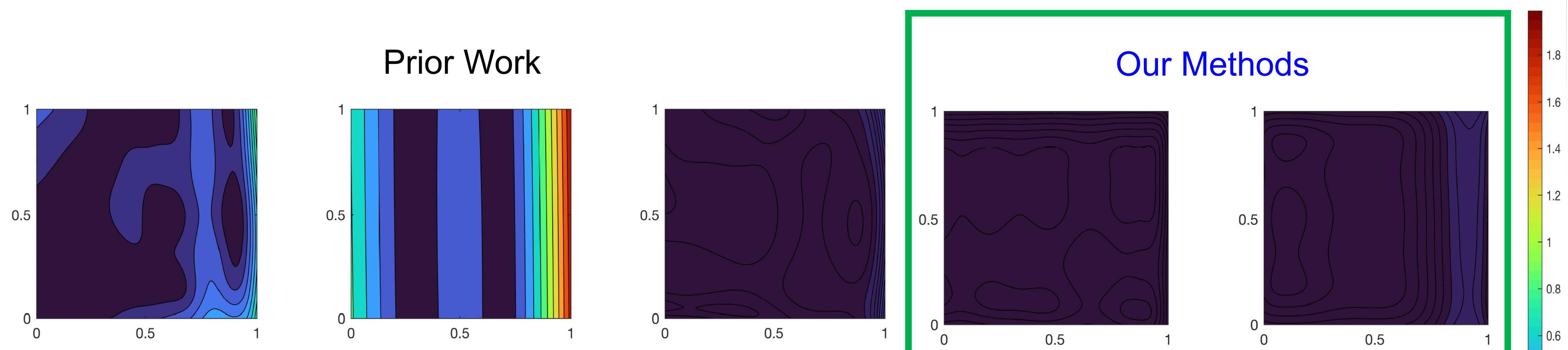

$$\boldsymbol{z}_k = \boldsymbol{W}_k\boldsymbol{x} + \boldsymbol{b}_k$$
$$\text{M-MGN}(\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{x} + \sum_{k=1}^{K} s_k(\boldsymbol{z}_k)\,\boldsymbol{W}_k^\top \sigma_k(\boldsymbol{z}_k)$$

Jacobian is PSD for $\sigma_\ell = \nabla s_\ell$ where $s_\ell(\cdot)$ is a twice differentiable, convex, and non-negative function

## GRADIENT FIELD RESULTS

We estimate the gradient field of convex $f(\boldsymbol{x})$ over the 2D unit square

$$f(\boldsymbol{x}) = x_1^4 + \frac{x_2}{2} + \frac{x_1 x_2}{2} + \frac{3x_2^2}{2} - \frac{x_2^3}{3}$$



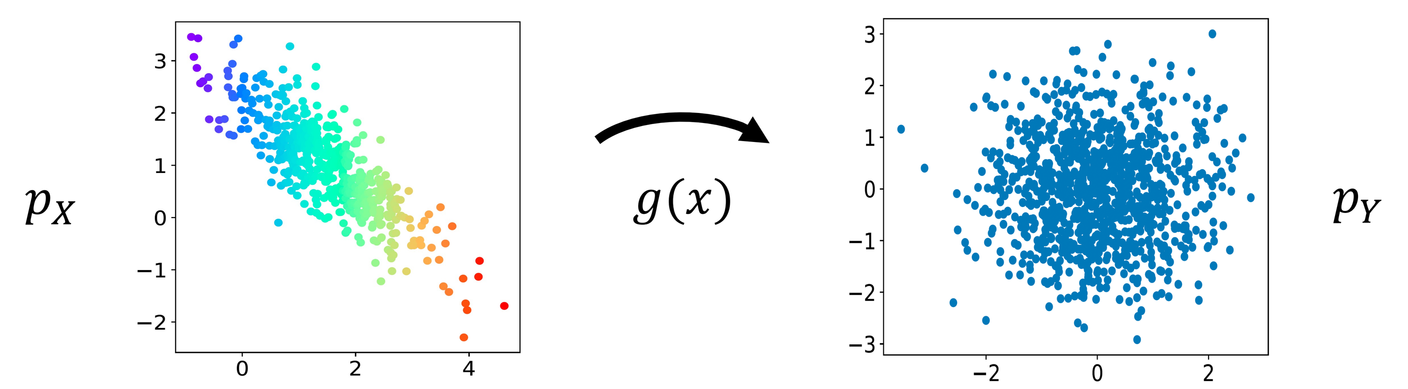

| ICGN [2] | ICNN [3] | ICNN [3] | C-MGN | M-MGN |
|---|---|---|---|---|
| -15.00 dB | -4.15 dB | -30.88 dB | **-39.10 dB** | -32.31 dB |
| 15 Parameters | 78 Parameters | 163 Parameters | **14 Parameters** | 22 Parameters |

Plots show squared Euclidean error over 2D unit square. Mean squared error given in dB.

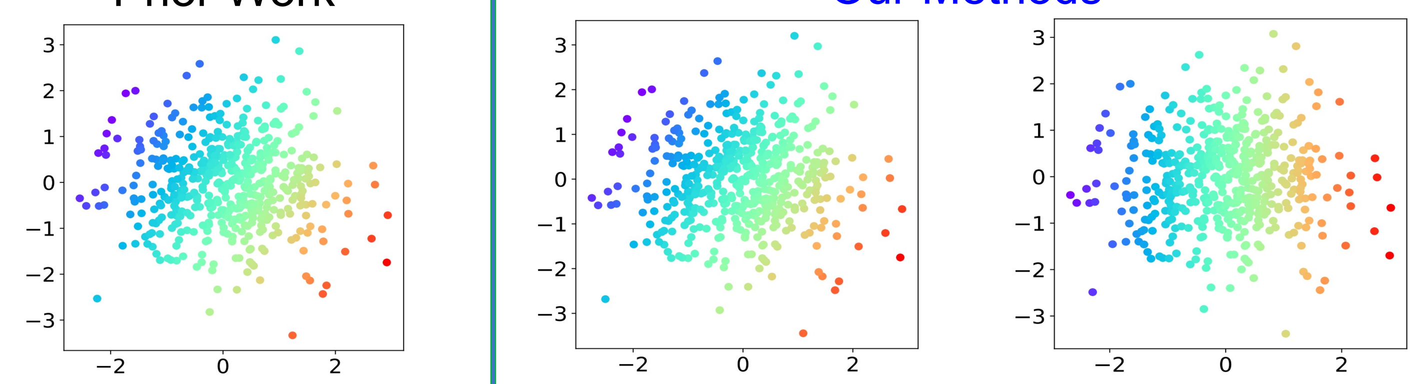
## OPTIMAL TRANSPORT RESULTS

We parameterize $g(x)$ as a neural network and minimize Negative Log Likelihood (NLL) to solve Optimal Transport

$$\inf_g -\log\left(p_Y(g(x))\right) \qquad \inf_{g: g(x)\sim p_Y} \mathbb{E}_{x\sim p_X}||x - g(x)||_2^2$$




Source Distribution | Standard Gaussian Target Distribution



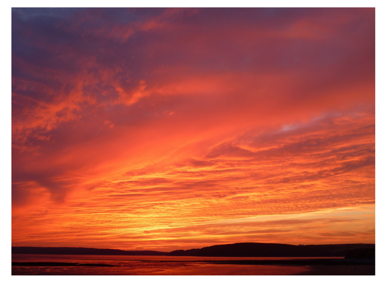

| CP Flow [4] | C-MGN | M-MGN |
|---|---|---|
| NLL: 3.10 | NLL: 2.86 | **NLL: 2.87** |
| Cost: 3.47 | Cost: 3.30 | **Cost: 3.21** |

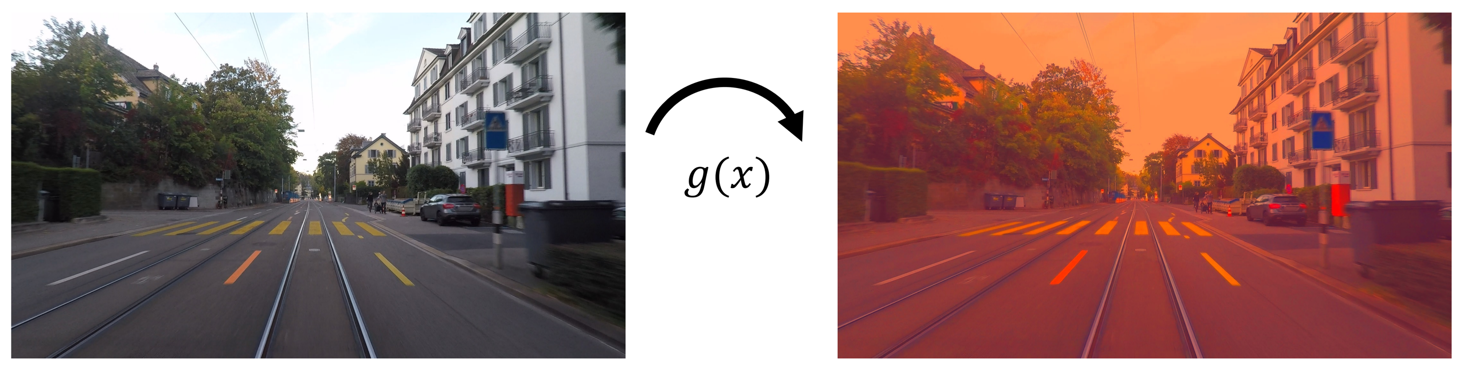
*Applying Optimal Transport to Autonomous Driving Data*

Efficiently generating labeled data for Domain Adaptation problems

We apply MGNs to map road images in the Dark Zurich Dataset [5] to new lighting conditions

Target Image of Sunset (Desired Color Distribution $p_Y$):






Source Image [5] (Source Color Distribution $p_X$) | M-MGN Mapped Image with Target Color Distribution

## Selected References

1. Yann Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on Pure and Applied Mathematics*, vol. 44, no. 4, pp. 375–417, 1991
2. Jack Richter-Powell, Jonathan Lorraine, and Brandon Amos, "Input convex gradient networks," *arXiv* preprint arXiv:2111.12187, 2021.
3. Brandon Amos, Lei Xu, and J Zico Kolter, "Input convex neural networks," in *International Conference on Machine Learning. PMLR*, 2017, pp. 146–155.
4. Chin-Wei Huang, Ricky T. Q. Chen, Christos Tsirigotis, and Aaron Courville, "Convex potential flows: Universal probability distributions with optimal transport and convex optimization," in *International Conference on Learning Representations*, 2021
5. Christos Sakaridis, Dengxin Dai, and Luc Van Gool, "Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation," in *The IEEE International Conference on Computer Vision (ICCV)*, 2019.