# Robust Stereo Matching by Risk Minimization (ICLR 2024)

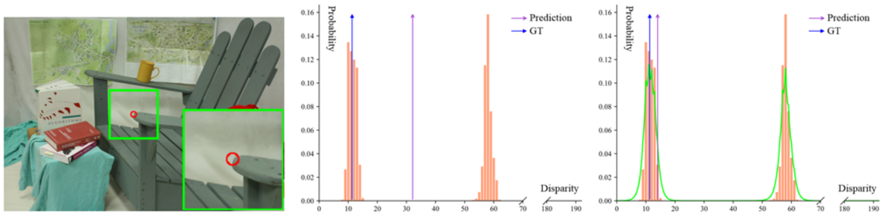
2023.11.9

# Contribution & Motivation

1. 解决立体匹配中连续视差建模的重要性。我们通过将视差预测问题视为寻找视差值最小风险的搜索问题。我们证明了常用的视差期望是所提出的风险公式中 L2 误差函数的一个特例，它对多模态分布敏感，可能导致解过于平滑。相反，我们主张在风险最小化过程中使用 L1 误差函数。

2.在本文中，我们通过计算所提出的风险函数的导数并执行其连续优化来寻找解决方案。通过对视差分类分布进行插值，我们定义了连续概率密度函数。然后，我们提出了一种二分搜索算法来找到有效地最小化所提出的风险的最佳视差。为了实现端到端网络训练，我们通过隐函数定理（计算最终视差相对于分类分布的后向梯度。

## Method: Overview of Network




(a) Image (b) Expectation (c) Ours

$$y = \int x p(x; \mathbf{p}) dx.$$

$$\texttt{argmin}_y F(y, \mathbf{p}) = \int x p(x; \mathbf{p}) dx \text{ when } \mathcal{L}(y, x) = (y - x)^2$$

$$y = \texttt{argmin}_y F(y, \mathbf{p}) = \texttt{argmin}_y \int |y - x| p(x; \mathbf{p}) dx.$$

$$G(y, \mathbf{p}) \triangleq \frac{\partial F(y, \mathbf{p})}{\partial y} = \sum_i p_i \texttt{Sign}(y - d_i)(1 - \exp - \frac{|y - d_i|}{\sigma}) = 0$$

L(y,x)是y和x之间的误差函数。我们所说的风险是指，如果我们将 y 作为预测差异，相对于真实情况会有多少误差。由于在进行预测时无法获得准确的基本事实，因此我们将所有可能的基本事实差异的误差与分布 p(x; p) 进行平均。

然而，众所周知，L2 范数并不稳健，并且容易出现异常值。如图2(b)所示，当存在多种模式时，期望不准确，求导就完了。

first-order derivative is a nondecreasing function

# Method: Overview of Network

**Algorithm 1** Forward Prediction

**Require:** $\tau > 0$, $\sigma > 0$, $\mathbf{d} = [d_1, ..., d_N]$, $d_1 < d_2 < \cdots < d_N$, and $\mathbf{p} = [p_1, ..., p_N]$

```
d^l ← d_1                                              ▷ Initialize search boundaries
d^r ← d_N
g ← τ + 1                                              ▷ Initialize the derivative
while |g| > τ do
    d^m ← (d^l + d^r)/2.0                              ▷ Compute the mid point
    g ← Σ_i p_i Sign(d^m − d_i)(1 − exp −|d^m − d_i|/σ) ▷ Compute the derivative by Eq.(5)
    if g > 0 then                                      ▷ Update search boundaries
        d^r ← d^m
    else
        d^l ← d^m
    end if
end while
return d^m                                             ▷ Return the mid point
```

n

$$dG(y, \mathbf{p}) = \frac{\partial G}{\partial y} dy + \frac{\partial G}{\partial \mathbf{p}} d\mathbf{p} = 0.$$

ing the terms, we obtain

$$\frac{dy}{d\mathbf{p}} = -\frac{\partial G / \partial \mathbf{p}}{\partial G / \partial y} = [\ldots, \frac{\sigma \mathtt{Sign}(d_i - y)(1 - \exp - \frac{|y - d_i|}{\sigma})}{\sum_j p_j \exp - \frac{|y - d_j|}{\sigma}}, \ldots]^T.$$



微分

# Experiment

| Method | Param (M) | Time (s) | > 2px Noc | > 2px All | > 3px Noc | > 3px All |
|---|---|---|---|---|---|---|
| LEAStereo (Cheng et al., 2020) | 1.81 | | 1.90 | 2.39 | 1.13 | 1.45 |
| CFNet (Shen et al., 2021) | 21.98 | 0.12 | 1.90 | 2.43 | 1.23 | 1.58 |
| ACVNet (Xu et al., 2022) | 6.84 | 0.15 | 1.83 | 2.34 | 1.13 | 1.47 |
| ACFNet (Chen et al., 2021) | | | 1.83 | 2.35 | 1.17 | 1.54 |
| NLCA-Net v2 (Rao et al., 2022) | | | 1.83 | 2.34 | 1.11 | 1.46 |
| CAL-Net (Chen et al., 2021) | | | 1.74 | 2.24 | 1.19 | 1.53 |
| CREStereo (Li et al., 2022) † | | | 1.72 | 2.18 | 1.14 | 1.46 |
| LaC+GANet (Liu et al., 2022a) | 9.43 | | 1.72 | 2.26 | 1.05 | 1.42 |
| IGEV (Xu et al., 2023)† | 12.60 | 0.32 | 1.71 | 2.17 | 1.12 | 1.44 |
| PCWNet (Shen et al., 2022) | 34.27 | 0.23 | 1.69 | 2.18 | 1.04 | 1.37 |
| **Ours** | **11.96** | **0.32** | **1.58** | **2.20** | **1.00** | **1.44** |

| Method | Param (M) | Time (s) | All D1_bg | All D1_fg | All D1_all | Noc D1_bg | Noc D1_fg | Noc D1_all |
|---|---|---|---|---|---|---|---|---|
| LEAStereo (Cheng et al., 2020) | 1.81 | | 1.40 | 2.91 | 1.65 | 1.29 | 2.65 | 1.51 |
| CFNet (Shen et al., 2021) | 21.98 | 0.12 | 1.54 | 3.56 | 1.88 | 1.43 | 3.25 | 1.73 |
| ACVNet (Xu et al., 2022) | 6.84 | 0.15 | 1.37 | 3.07 | 1.65 | 1.26 | 2.84 | 1.52 |
| ACFNet (Chen et al., 2021) | | | 1.51 | 3.80 | 1.89 | 1.36 | 3.49 | 1.72 |
| NLCA-Net v2 (Rao et al., 2022) | | | 1.41 | 3.56 | 1.77 | 1.28 | 3.22 | 1.60 |
| CAL-Net (Chen et al., 2021) | | | 1.59 | 3.76 | 1.95 | 1.45 | 3.42 | 1.77 |
| CREStereo (Li et al., 2022) † | | | 1.45 | 2.86 | 1.69 | 1.33 | 2.60 | 1.54 |
| LaC+GANet (Liu et al., 2022a) | 9.43 | | 1.44 | 2.83 | 1.67 | 1.26 | 2.64 | 1.49 |
| IGEV (Xu et al., 2023) † | 12.60 | 0.32 | 1.38 | 2.67 | 1.59 | 1.27 | 2.62 | 1.49 |
| DLNR (Zhao et al., 2023) | 54.72 | 0.39 | 1.60 | 2.59 | 1.76 | 1.45 | 2.39 | 1.61 |
| PCWNet (Shen et al., 2022) | 34.27 | 0.23 | 1.37 | 3.16 | 1.67 | 1.26 | 2.93 | 1.53 |
| CroCo-Stereo (Weinzaepfel et al., 2023)† | 417.15 | | 1.38 | 2.65 | 1.59 | 1.30 | 2.56 | 1.51 |
| **Ours** | **11.96** | **0.32** | **1.40** | **2.76** | **1.63** | **1.25** | **2.62** | **1.48** |

KITTI 2012 2015

# Ablation

Table 8: Ablation studies on Middlebury training set of quarter resolution. The first and second bests are in red and blue respectively. **Our method** in bold.

| Backbone | Training | Test | Param (M) | Time (s) | > 1px | | > 2px | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Noc | All | Noc | All |
| ACVNet(Xu et al., 2022) | Expectation | Expectation | 6.84 | 0.12 | 22.68 | 26.49 | 13.54 | 16.49 |
| | Expectation | L1-Risk | 6.84 | 0.18 | 22.32 | 26.14 | 13.13 | 16.05 |
| PCWNet(Shen et al., 2022) | Expectation | Expectation | 34.27 | 0.19 | 16.80 | 21.36 | 8.93 | 12.62 |
| | Expectation | L1-Risk | 34.27 | 0.26 | 16.53 | 21.08 | 8.65 | 12.30 |
| **Ours** | Expectation | Expectation | 11.96 | 0.17 | 9.88 | 13.27 | 4.92 | 7.29 |
| | Expectation | L1-Risk | 11.96 | 0.25 | 9.83 | 13.22 | 4.90 | 7.27 |
| | L1-Risk | Expectation | 11.96 | 0.17 | 9.83 | 13.19 | 4.79 | 7.06 |
| | **L1-Risk** | **L1-Risk** | **11.96** | **0.25** | **9.32** | **12.63** | **4.49** | **6.70** |

36 manydepth
37 monorec

# Thanks