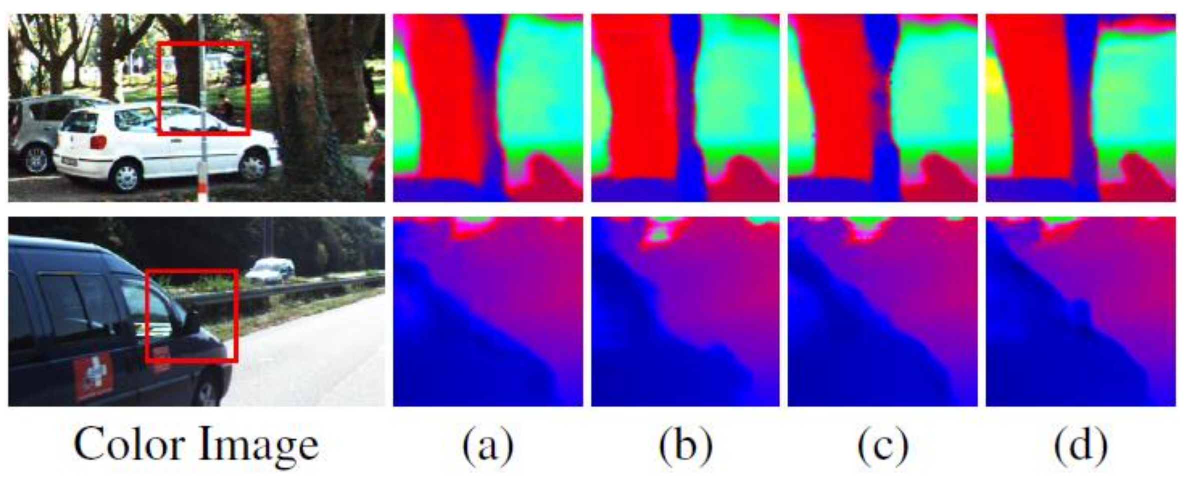
国际人工智能会议

AAAI 2021 论文北京预讲会

FCFR-Net: Feature Fusion based Coarse-to-Fine Residual Learning for Monocular Depth Completion

Lina Liu, Xibin Song, Xiaoyang Lyu, Junwei Diao, Mengmeng Wang, Yong Liu
and Liangjun Zhang



(a) STD
(b) Deep Lidar
(c) CSPN++
(d) Our

1. Motivation:

Information fusion is insufficient

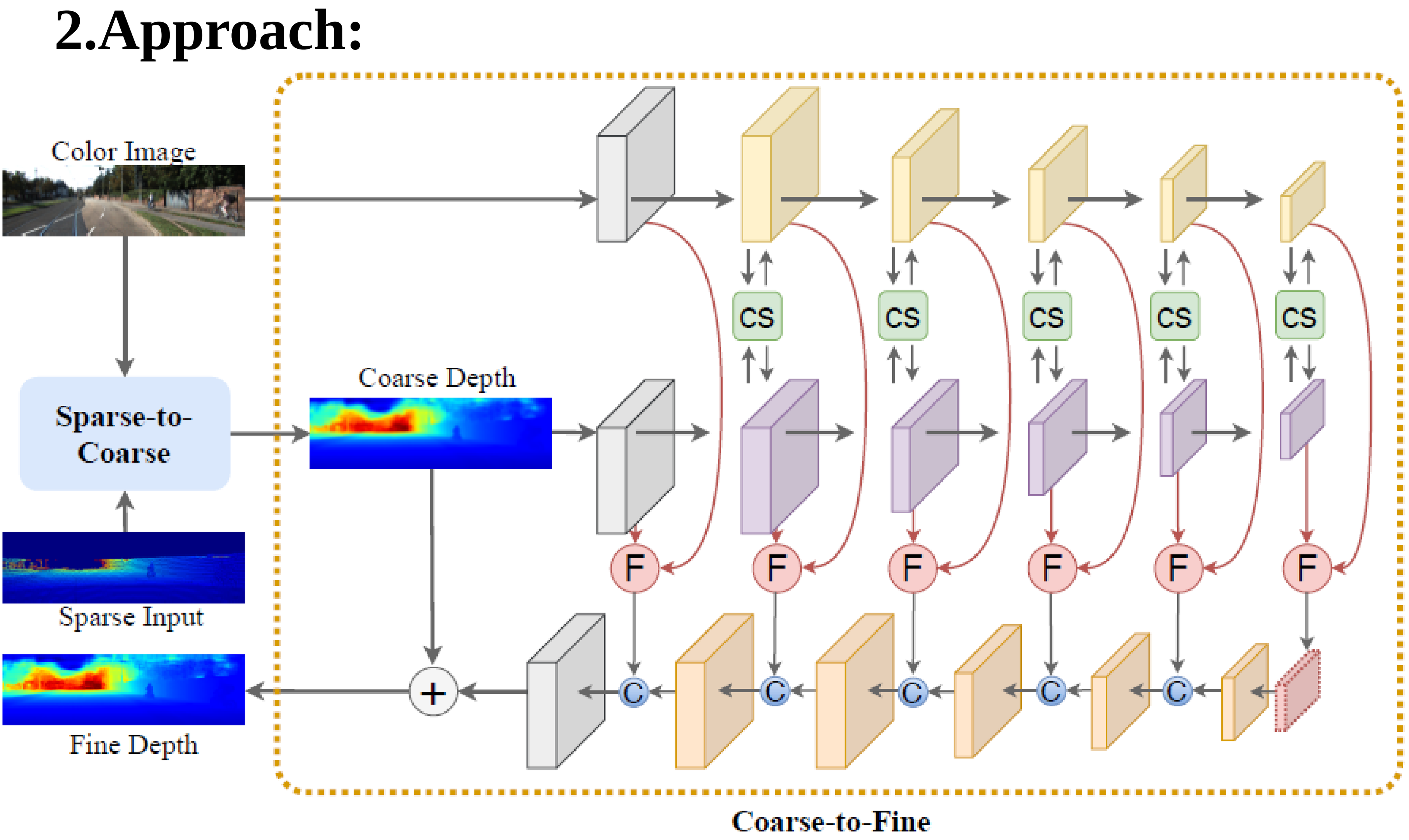
Contributions:

(1). Formulate depth completion as a two stage task, and design a coarse-to-fine residual learning based framework;

(2). Design channel shuffle extraction operation, which effectively fuses the features of color and depth information at the multi-scale feature levels;

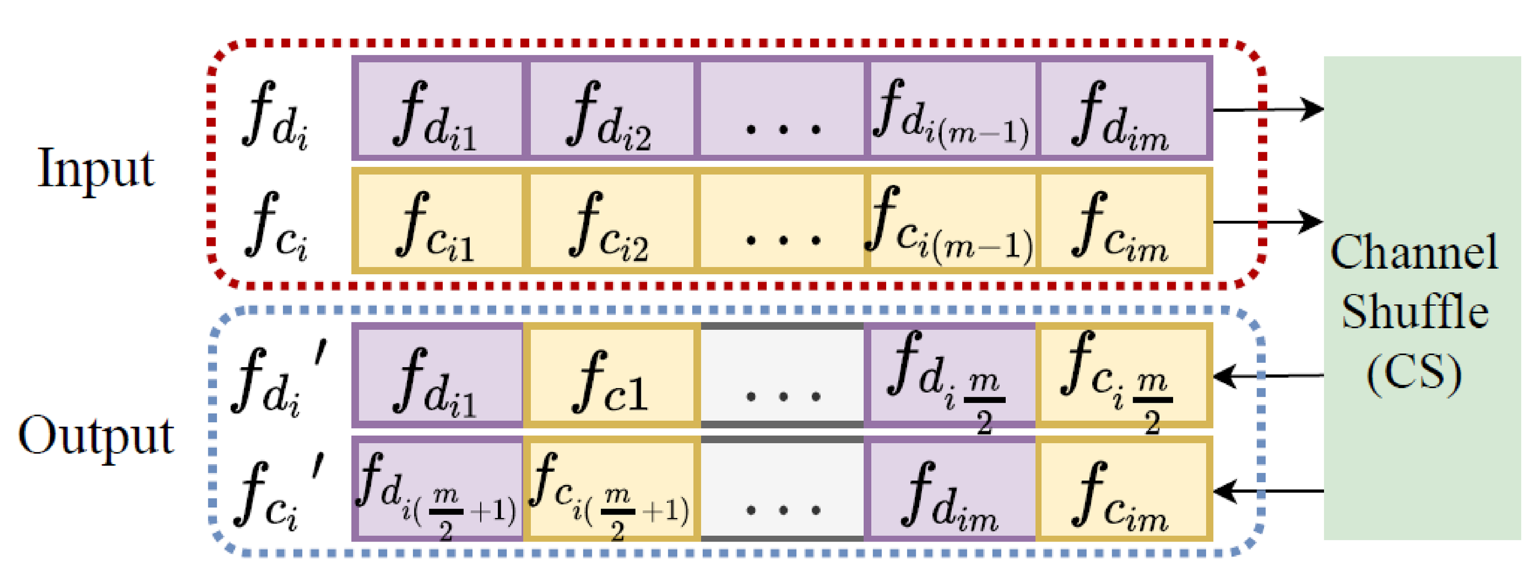
(3). A energy based fusion operation is utilized to further sufficiently fuse the features obtained by channel shuffle extraction.

2.Approach:



Overview of network architecture

2.1 channel shuffle



Given depth and color features of the i-th convolution block, $f_{d_i} = \{f_{d_{i1}}, \ldots, f_{d_{iM}}\}$, $f_{c_i} = \{f_{c_{i1}}, \ldots, f_{c_{iM}}\}$, where M is the number of channels, the output of channel shuffle are:

$$f'_{d_i} = \{f_{d_{i1}}, f_{c_{i1}}, \ldots, f_{d_{i\frac{M}{2}}}, f_{c_{i\frac{M}{2}}}\}$$

$$f'_{c_i} = \{f_{d_{i\frac{M}{2}+1}}, f_{c_{i\frac{M}{2}+1}}, \ldots, f_{d_{iM}}, f_{c_{iM}}\}$$

2.2 Energy based fusion

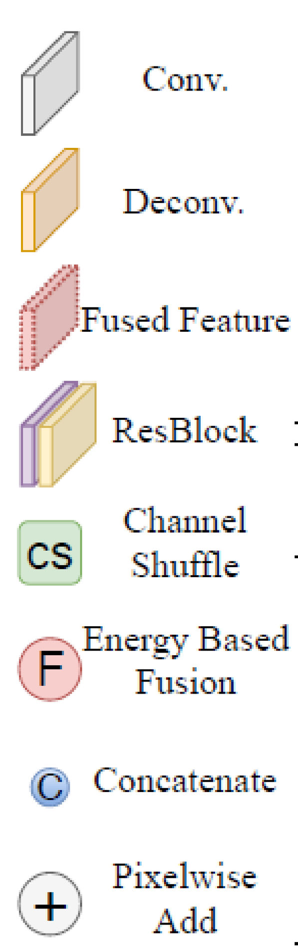
Suppose that H, W are the height and width of a feature map f_{ij}, where $i \in [0, N], j \in [1, M]$, N is number of feature and M is the number of channels. $f_{k_{ij}}(m,n)$ is the feature value at (m, n), where $m \in [1, H], n \in [1, W]$, and f_k represents color and depth features. $E_k(m,n)$ means the energy in region $L \times L$ centered at (m, n). $k \in [1, 2]$ mean color and depth information.

$$E_{k_{ij}}(m,n) = \sum_{a=-m'}^{m'} \sum_{b=-n'}^{n'} \omega(f_{k_{ij}}(m+a, n+b))^2$$

$$f_{o_{ij}}(m,n) = \begin{cases} \sigma f_{1_{ij}}(m,n), & E_{1_{ij}}(m,n) \geq E_{2_{ij}}(m,n) \\ \sigma f_{2_{ij}}(m,n), & E_{1_{ij}}(m,n) < E_{2_{ij}}(m,n) \end{cases}$$

where σ is the coefficient, and f_o is the output.

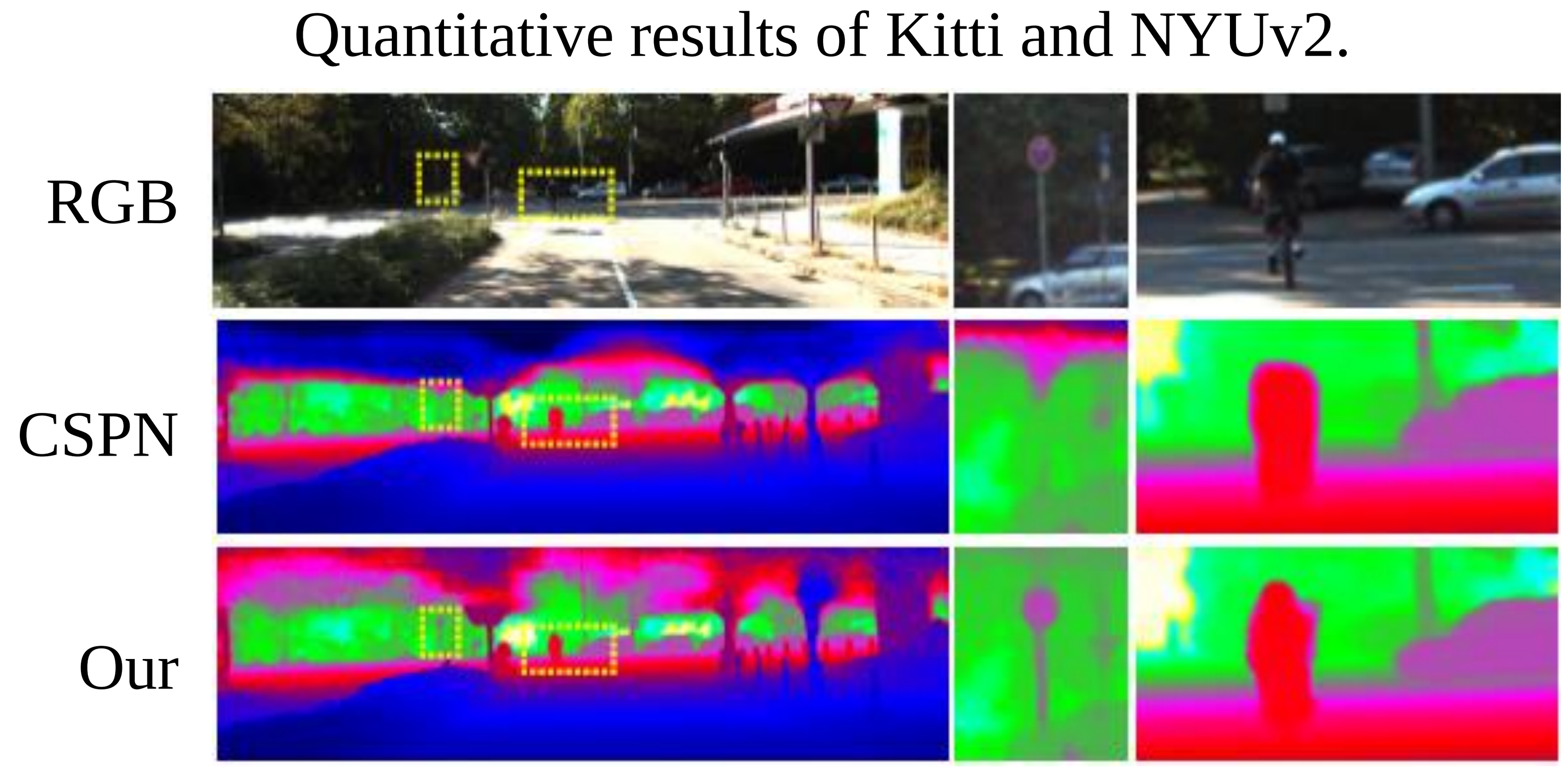
3. Experiment



Method	RMSE mm	MAE mm	iRMSE 1/km	iMAE 1/km
CSPN	1019.64	279.46	2.93	1.15
STD	814.73	249.95	2.80	1.21
CG (Lee et al. 2020)	807.42	253.98	2.73	1.33
RV	792.80	225.81	2.42	0.99
PwP (Xu et al. 2019)	777.05	235.17	2.42	1.13
RGBG&C	772.87	215.02	2.19	0.93
MSG-CHN (Li et al. 2020)	762.19	220.41	2.30	0.98
DeepLiDAR (Qiu et al. 2019)	758.38	226.50	2.56	1.15
Uber (Chen et al. 2019)	752.88	221.19	2.34	1.14
CSPN++ (Cheng et al. 2020)	743.69	209.28	2.07	0.90
NLSPN (Park et al. 2020)	741.68	**199.59**	**1.99**	**0.84**
Ours	**735.81**	217.15	2.20	0.98

Method	RMSE m	REL m	$\delta_{1.25}$	$\delta_{1.25^2}$	$\delta_{1.25^3}$
STD_18	0.230	0.044	97.1	99.4	99.8
Sparse-to-Coarse	0.123	0.026	99.1	**99.9**	**100.0**
CSPN	0.117	0.016	99.2	**99.9**	**100.0**
CSPN++ (Cheng et al. 2020)	0.116	-	-	-	-
DeepLiDAR (Qiu et al. 2019)	0.115	0.022	99.3	**99.9**	**100.0**
PwP (Xu et al. 2019)	0.112	0.018	**99.5**	**99.9**	**100.0**
Ours	**0.106**	**0.015**	**99.5**	**99.9**	**100.0**

Quantitative results of Kitti and NYUv2.



Qualitative results of Kitti.

Contact:

inter position for 3D vision and robotics (RAL of Baidu Research)

Email: ral_jobs@baidu.com; songxibin@baidu.com

主办方：中国中文信息学会青年工作委员会

承办方：智源社区