# BFC-BL: Few-Shot Classification and Segmentation- combining Bi-directional Feature Correlation and Boundary constraint

HaiBiao Yang, Bi Zeng, PengFei Wei*, JianQi Liu

Guangdong University of Technology

* Corresponding author

Code and data are publicly available: https://github.com/XIAO1HAI/BFC-BL

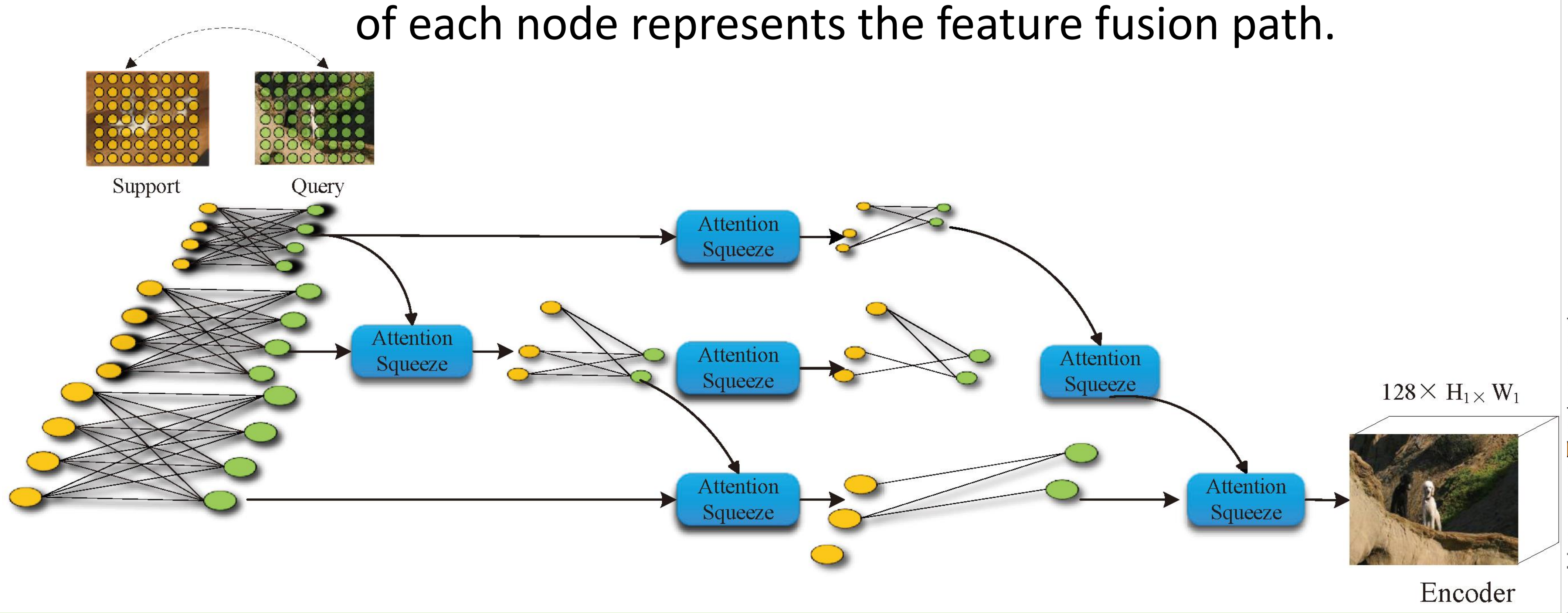
## Few-Shot Classification and Segmentation

Our contributions:

- (1) This paper proposes bidirectional feature fusion for few-shot classification and segmentation. The cross-scale bidirectional semantic correlation fusion module (BFCP) integrates deep semantic and shallow spatial correlation measurements, facilitating rapid learning of the correlation between support and query samples.
- (2) To enhance the segmentation performance of the model, we introduce a boundary constraint function, which is integrated with the region loss to construct a multi-level weight ratio loss function. This loss function guides the model to learn boundary information, thereby improving its ability to accurately segment target boundary pixels.
- (3) By leveraging the two-way high-level semantic correlation between support and query samples, low-level spatial correlation, and the boundary constraints, we construct an end-to-end few-shot classification and segmentation network model. This model enables rapid and effective few-shot classification and segmentation, providing significant advancements in the field.

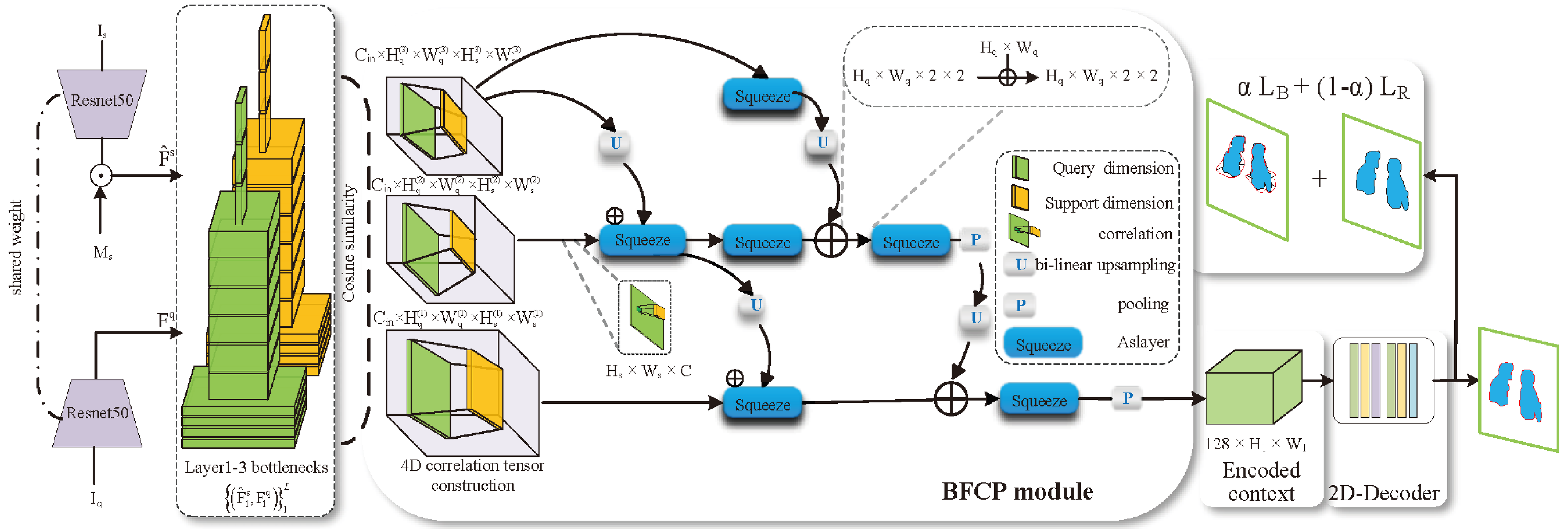
## Bidirectional Feature Correlation fusion for Multi-level

The green node represents the feature map of each layer of feature extraction, the blue node represents the feature processing layer (such as the attention layer), and the line of each node represents the feature fusion path.




Acknowledgements This work was supported in part by the Key Technology Project of Shunde District, Foshan City, under Grant 2130218003002.

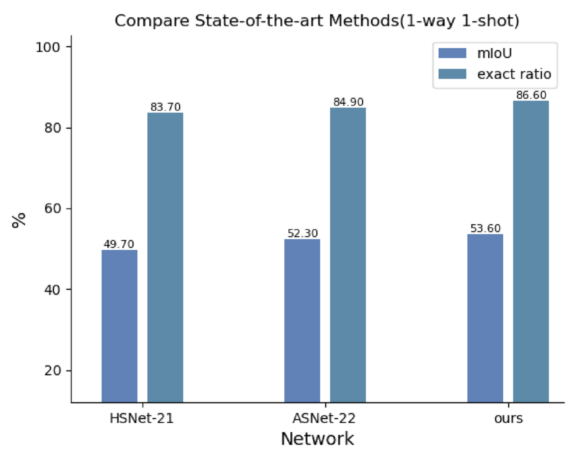
## Overall architecture of the BFC-BL model




- ResNet50 is used as the backbone network for feature extraction. This process involves computing the correlation between the query and support samples, resulting in a 4D tensor constructed using cosine similarity.
- (2) A cross-scale bidirectional feature correlation fusion module (BFCP) is designed and embedded into the encoder structure to perform an interactive fusion of deep semantic correlation and shallow spatial correlation features.
- (3) Utilizing the context information obtained from the encoder, a 2D decoder decodes the learned encoder knowledge and generates the final output.

## Compare the role of individual modules and Experimental Results

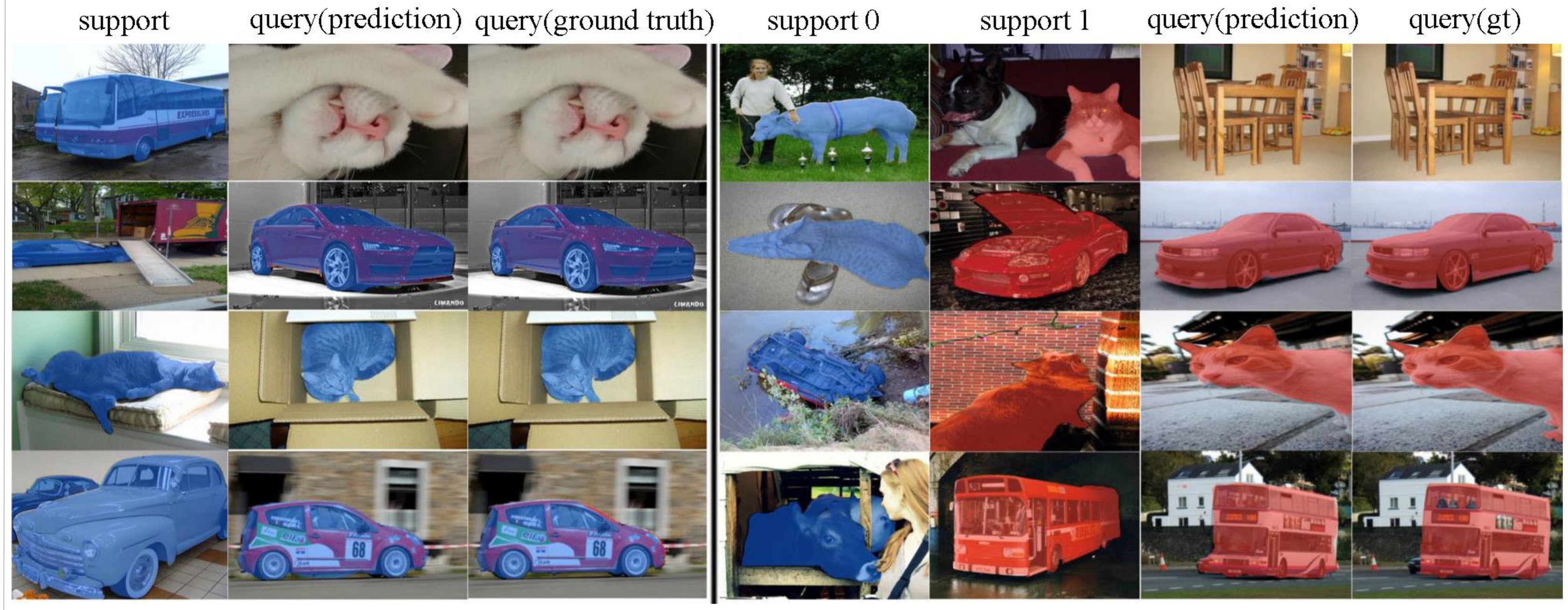
Uses the 0/1 accuracy rate exact ratio $ER = 1[y_{gt} = y_c]$ and union ratio $mIoU = \frac{1}{C}\sum_c IoU_c$ evaluate.



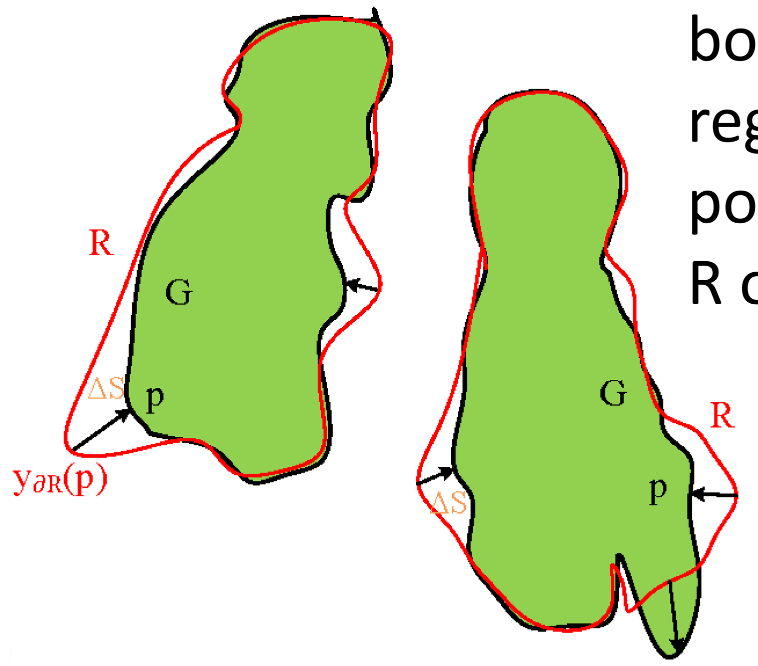

| Methods | 1-way 1-shot | | | | | | | | | | 2-way 1-shot | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | classification 0/1 exact ratio (%) | | | | | Segmentation mIoU (%) | | | | | classification 0/1 exact ratio (%) | | | | | Segmentation mIoU (%) | | | | |
| | $5^0$ | $5^1$ | $5^2$ | $5^3$ | avg. | $5^0$ | $5^1$ | $5^2$ | $5^3$ | avg. | $5^0$ | $5^1$ | $5^2$ | $5^3$ | avg. | $5^0$ | $5^1$ | $5^2$ | $5^3$ | avg. |
| PANet[22] | 69.9 | 67.7 | 68.8 | 69.4 | 69.0 | 32.8 | 45.8 | 31.0 | 35.1 | 36.2 | 56.2 | 47.5 | 44.6 | 55.4 | 50.9 | 33.3 | 46.0 | 31.2 | 38.4 | 37.2 |
| PFENet[20] | 69.8 | 82.4 | 68.1 | 77.9 | 74.6 | 38.3 | 54.7 | 35.1 | 43.8 | 43.0 | 22.5 | 61.7 | 40.3 | 39.5 | 41.0 | 31.1 | 47.3 | 30.8 | 32.2 | 35.3 |
| HSNet[12] | 86.6 | 84.8 | 76.9 | 86.3 | 83.7 | 49.1 | 59.7 | 41.0 | 49.0 | 49.7 | 68.0 | 73.2 | 57.0 | 70.9 | 67.3 | 42.4 | 53.7 | 34.0 | 43.9 | 43.5 |
| ASNet[8] | 84.9 | **89.6** | 79.0 | 86.2 | 84.9 | 51.7 | 61.5 | 43.3 | 52.8 | 52.3 | 68.5 | 76.2 | 58.6 | 70.0 | 68.3 | 48.5 | 58.3 | 36.3 | 48.3 | 47.8 |
| ours | **87.4** | 89.4 | **81.1** | **88.5** | **86.6** | **52.9** | **62.7** | **44.6** | **54.2** | **53.6** | **70.6** | **77.0** | **60.4** | **72.1** | **70.1** | **50.2** | **59.2** | **37.8** | **49.7** | **49.2** |

| BFCP module | boundary constraints | classification 0/1 exact ratio (%) | Segmentation mIoU (%) |
|---|---|---|---|
| ✓ | ✘ | $86.1_{+1.2}$ | $52.5_{+0.2}$ |
| ✘ | ✓ | $85.3_{-0.8}$ | $53.2_{+0.9}$ |
| ✓ | ✓ | $86.6_{+1.7}$ | $53.6_{+1.3}$ |

## Segmentation Result




## Construction of Boundary Constraints




∂G is the set of all points on the boundary of the ground truth region G, ∂R is the set of boundary points of the segmentation region R output by the network model.

$$\int_p^{y_{\partial R}(p)} 2D_G(q)dq = \int_0^{\|y_{\partial R}(p)-p\|} 2D_G dD_G = \|y_{\partial R}(p) - p\|^2.$$

## Accelerated rate of convergence

The accuracy and mIoU of our proposed model converge between 200-250 epochs, and the baseline requires 500 epochs.