ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10448447

LEARNING SPECTRAL CANONICAL $\mathcal{F}$-CORRELATION REPRESENTATION FOR FACE SUPER-RESOLUTION

Yun-Hao Yuan[⋆,†], *Mingzhi Hao*[⋆], *Yun Li*[⋆], *Jipeng Qiang*[⋆], *Yi Zhu*[⋆], *Xiaobo Shen*[‡]

[⋆]School of Information Engineering, Yangzhou University, Yangzhou, China
[†]School of Computer Science, Fudan University, Shanghai, China
[‡]School of Computer Science, Nanjing University of Science and Technology, Nanjing, China

ABSTRACT

Face super-resolution (FSR) is a powerful technique for restoring high-resolution face images from the captured low-resolution ones with the assistance of prior information. Existing FSR methods based on explicit or implicit covariance matrices are difficult to reveal complex nonlinear relationships between features, as conventional covariance computation is essentially a linear operation process. Besides, the limited number of training samples and noise disturbance lead to the deviation of sample covariance matrices. To solve these issues, we propose a novel FSR method via using spectral canonical $\mathcal{F}$-correlation representation. The proposed method first defines intra-resolution and inter-resolution covariation matrices by considering the nonlinear relationship between different features, and then uses the fractional order idea to rebuild covariation matrices. The qualitative and quantitative results have validated the superiority of the proposed method.

Index Terms— face super-resolution, canonical correlation analysis, fractional order, nonlinear feature relationship

1. INTRODUCTION

Face super-resolution (FSR) can be seen as a domain-specific image super-resolution (SR) problem, which infers the high-resolution (HR) face images from the low-resolution (LR) ones via considering unique priors on the face. It has been widely used in various face analysis-related applications, such as face detection [1], face recognition [2], face re-identification [3], and so on. Because of the highly under-determined constraints and possible data noise, FSR is a seriously ill-posed problem, thus making it much challenging. In recent years, learning-based FSR has achieved great progress, which tries to take advantage of machine learning technologies to learn the complicated relationship between HR and LR face data via given training set. In this work, we focus on this type of FSR techniques.

Thus far, there have been a lots of useful learning-based FSR techniques to be developed; see, for instance, [4, 5, 6, 7]. Among these FSR methods, an attractive research direction is to explore and exploit the intrinsic association information of LR and HR face images during super-resolution process. Undoubtedly, canonical correlation analysis (CCA) [8], a well-known statistical tool for measuring the linear correlation between two sets of variables, is one of the most appropriate tools for modeling the correlation between LR and HR faces. So far, many CCA-related methods have been applied to FSR tasks. For instance, in 2010, Huang et al. [9] pioneered the application of CCA to FSR and proposed a coherent local linear reconstruction SR (CLLR-SR) method, where the correlation between LR and HR images is maximized. Later, An et al. [10] proposed a 2D CCA-based FSR method directly based on 2D face images rather than face vectors.

However, the preceding CCA-related FSR methods are based on covariance matrices in nature, which only model the linear relation between different features. Moreover, these methods must face the singularity of within-set covariance matrices when the facial feature dimension is larger than the number of training faces. In this paper, we attempt to borrow the idea from our previous work [11] to overcome these two intractable problems in FSR. On the other hand, in realistic scenario, it is well-known that intra- and inter-set sample covariation matrices in [11] might deviate from the true ones due to the noise and limited number of LR/HR training face images. Existing studies [12, 13] in the filed of feature extraction and reduction have shown that spectral reconstruction can improve the problem of this deviation, obtaining encouraging learning performance.

Based on the above-mentioned considerations, in this paper, we propose a novel spectral canonical $\mathcal{F}$-correlation representation method for face hallucination, dubbed SCFR-FSR, where the spectra of the sample covariation matrix are reconstructed via using fractional order strategy. Our proposed method can not only model the nonlinear relationship between LR and HR facial features, but also alleviates the

This work is supported by the China Postdoctoral Science Foundation under grant 2020M670995, the National Natural Science Foundation of China under grants 62176126, 62076217, and 61906060, and the Natural Science Foundation of Jiangsu Province of China under grant BK20190440.

deviation problem of sample covariation matrices. Experimental results show that our proposed SCFR-FSR method is more encouraging than existing methods.

2. PROPOSED METHOD

2.1. Formulation

Assume that LR and HR training face images are $L = [l_1, l_2, \cdots, l_n] \in R^{p\times n}$ and $H = [h_1, h_2, \cdots, h_n] \in R^{q\times n}$, where n is the number of training images, p and q separately denote the dimension of LR and HR image vectors. To reduce the data noise and computational burden, we use principal component analysis (PCA) to extract the useful features after centralizing every HR/LR face image:

$$\begin{aligned} \tilde{H} &= P_h^T \left(H - \mu_h \mathbf{1}^T\right) \in R^{d_h\times n}, \\ \tilde{L} &= P_l^T \left(L - \mu_l \mathbf{1}^T\right) \in R^{d_l\times n}, \end{aligned} \tag{1}$$

where $\mu_h \in R^q$ and $\mu_l \in R^p$ denote the average HR and LR faces, respectively, $\mathbf{1} \in R^n$ is the vector of all ones, $P_h \in R^{q\times d_h}$ and $P_l \in R^{p\times d_l}$ are separately the PCA transformations, $d_h \le \min(q, n)$ and $d_l \le \min(p, n)$.

Rewriting $\tilde{H}$ and $\tilde{L}$ in the form of feature vectors, we are able to obtain the corresponding HR and LR facial feature matrices, i.e.,

$$\begin{aligned} \tilde{H} &= [f_h^1, f_h^2, ..., f_h^{d_h}]^T, \\ \tilde{L} &= [f_l^1, f_l^2, ..., f_l^{d_l}]^T, \end{aligned}$$

where $f_a^{t_a} \in R^n$ with $a \in \{h, l\}$ and $t_a = 1, 2, \cdots, d_a$. Assume that there is a nonlinear mapping $\psi(\cdot)$, which maps each feature vector $f_a^{t_a}$ to a new space via

$$\psi(f_a^{t_a}) = \left[\psi^1(f_a^{t_a}), \psi^2(f_a^{t_a}), \cdots, \psi^N(f_a^{t_a})\right]^T \tag{2}$$

where N is the dimension of ψ-generated new space. In this paper, $\psi(\cdot)$ is chosen as a kernel mapping associated with kernel function $ker(\cdot, \cdot)$. We make use of (2) to define $\mathcal{F}$-intra-resolution and $\mathcal{F}$-inter-resolution covariation matrices as:

$$K_{ab}(t_a, t_b) = \psi(f_a^{t_a})^T \psi(f_b^{t_b}) = ker(f_a^{t_a}, f_b^{t_b}), \tag{3}$$

where $K_{ab}(t_a, t_b)$ is the (t_a, t_b)-th element of matrix K_{ab}, $b \in \{h, l\}$, $t_b = 1, 2, \cdots, d_b$.

Via using (3), our SCFR-FSR method aims to seek pairs of projection directions $W_h \in R^{d_h\times d}$ and $W_l \in R^{d_l\times d}$ ($d \le \min(d_h, d_l)$) of HR and LR facial features by the following optimization problem:

$$\begin{aligned} &\max_{W_h, W_l} \ \mathrm{Tr}(W_h^T K_{hl}^{\beta} W_l) \\ &s.t. \quad W_h^T K_{hh}^{\alpha_h} W_h = I, \ W_l^T K_{ll}^{\alpha_l} W_l = I, \end{aligned} \tag{4}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, $I \in R^{d\times d}$ is the identity matrix, α_h, α_l, and β are the fractions, respectively, K_{hl}^{β} is computed by

$$K_{hl}^{\beta} = U\Sigma^{\beta}V^T, \ \ \Sigma^{\beta} = \mathrm{diag}(\sigma_1^{\beta}, \sigma_2^{\beta}, \cdots, \sigma_r^{\beta}) \tag{5}$$

with $0 \le \beta \le 1$, U and V are separately the left and right singular vector matrices of K_{hl}, $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$ are the descending nonzero singular values, $r = \mathrm{rank}(K_{hl})$, and $K_{aa}^{\alpha_a}$ ($a \in \{h, l\}$) is calculated by

$$K_{aa}^{\alpha_a} = Q_a \Gamma_a^{\alpha_a} Q_a^T, \ \ \Gamma_a^{\alpha_a} = \mathrm{diag}(\gamma_{a,1}^{\alpha_a}, \gamma_{a,2}^{\alpha_a}, \cdots, \gamma_{a,r_a}^{\alpha_a}) \tag{6}$$

with $0 \le \alpha_a \le 1$, Q_a is the eigenvector matrix of K_{aa}, $\gamma_{a,1} \ge \gamma_{a,2} \ge \cdots \ge \gamma_{a,r_a} > 0$ are the descending nonzero eigenvalues, and $r_a = \mathrm{rank}(K_{aa})$.

2.2. Optimization

For optimization problem (4), we solve it via the spectral decomposition method. Concretely, the first pair of projection directions, w_h and w_l, can be computed by:

$$\begin{aligned} &\max_{w_h, w_l} \ w_h^T K_{hl}^{\beta} w_l \\ &s.t. \quad w_h^T K_{hh}^{\alpha_h} w_h = 1, \ w_l^T K_{ll}^{\alpha_l} w_l = 1. \end{aligned} \tag{7}$$

The optimization problem (7) can be solved by the Lagrange multiplier technique, i.e.,

$$\begin{aligned} \mathcal{L}(w_h, w_l, \varphi_h, \varphi_l) = w_h^T K_{hl}^{\beta} w_l &- \frac{\varphi_h}{2}(w_h^T K_{hh}^{\alpha_h} w_h - 1) \\ &- \frac{\varphi_l}{2}(w_l^T K_{ll}^{\alpha_l} w_l - 1) \end{aligned} \tag{8}$$

with φ_h and φ_l as Lagrange multipliers. Let $\partial\mathcal{L}/\partial w_h = 0$ and $\partial\mathcal{L}/\partial w_l = 0$. Then, we get

$$\begin{aligned} (K_{hh}^{\alpha_h})^{-1} K_{hl}^{\beta} (K_{ll}^{\alpha_l})^{-1} K_{lh}^{\beta} w_h &= \varphi w_h, \\ (K_{ll}^{\alpha_l})^{-1} K_{lh}^{\beta} (K_{hh}^{\alpha_h})^{-1} K_{hl}^{\beta} w_l &= \varphi w_l, \end{aligned} \tag{9}$$

where $K_{lh}^{\beta} = (K_{hl}^{\beta})^T$ and φ ($= \varphi_h^2 = \varphi_l^2$) is the eigenvalue corresponding to the eigenvector w_h (or w_l). We are able to choose the first d pairs of eigenvectors $\{(w_h^i, w_l^i)\}_{i=1}^d$ of (9) to generate the projection matrices

$$\begin{aligned} W_h &= [w_h^1, w_h^2, \cdots, w_h^d], \\ W_l &= [w_l^1, w_l^2, \cdots, w_l^d]. \end{aligned}$$

2.3. Hallucinating Face

After obtaining W_h and W_l, we are able to calculate the correlation representations of PCA-transformed HR and LR features by the form of

$$\begin{aligned} H_{\mathrm{latent}} &= W_h^T \tilde{H} = \left[h_{\mathrm{latent}}^1, h_{\mathrm{latent}}^2, \cdots, h_{\mathrm{latent}}^n\right], \\ L_{\mathrm{latent}} &= W_l^T \tilde{L} = \left[l_{\mathrm{latent}}^1, l_{\mathrm{latent}}^2, \cdots, l_{\mathrm{latent}}^n\right]. \end{aligned}$$

Given a new LR face $l \in R^p$, its correlation feature can be obtained by

$$l_{\mathrm{latent}} = (P_l W_l)^T (l - \mu_l) \tag{10}$$

In L_{latent}, we select k nearest neighbors of l_{latent} to reconstruct it, assuming the index set of the k nearest neighbors is

id_k. Let $L^k_{\text{latent}} = \{l^k_{\text{latent}}\}_{k \in id_k}$. Then, l_{latent} can be reconstructed by the following optimization problem:

$$\min_{\eta} \left\| l_{\text{latent}} - L^k_{\text{latent}} \eta \right\|_2 \quad s.t. \quad \mathbf{1}^T \eta = 1, \tag{11}$$

where $\|\cdot\|_2$ denotes the 2-norm of a vector, $\eta \in R^k$, and $\mathbf{1} \in R^k$ is an all-ones vector. It is easy to find the optimal η of optimization problem (11). Let $H^k_{\text{latent}} = \{h^k_{\text{latent}}\}_{k \in id_k}$. Then, the corresponding HR correlation feature of l can be found by the form of

$$h_{\text{latent}} = H^k_{\text{latent}} \eta. \tag{12}$$

Using (12), we are able to obtain an initial HR face of l by the following

$$\text{initial}_h = P_h (W_h W_h^T)^{\dagger} W_h h_{\text{latent}} + \mu_h, \tag{13}$$

where $\dagger$ is the pseudo-inverse.

It should be noted that the reconstructed initial_h usually lacks facial detail information. Next, we introduce residual learning to refine the HR version in (13). To be specific, we first super-resolve each LR face of training set L using (13), obtaining the corresponding HR face image set Initial_H. Then, we separately compute HR and LR residual training face images by

$$\begin{aligned} \text{Residual}_H &= H - \text{Initial}_H, \\ \text{Residual}_L &= L - \text{Initial}_H \downarrow, \end{aligned} \tag{14}$$

where $\downarrow$ denotes a downsampling operation. In addition, we follow the patch idea [9] and divide all HR and LR residual training images into overlapping patches. Note that, we enlarge the size of LR residual faces to that of HR residual images before image partition. For HR and LR residual patches with the same position, we employ (4) to learn their correlation representations.

For a new LR face l, we first employ (13) to reconstruct the corresponding initial HR face initial_h, and then compute its LR residual face by

$$\text{residual}_l = l - \text{initial}_h \downarrow . \tag{15}$$

Likewise, we follow the preceding partition process to obtain the residual patches of l and use the above reconstruction process to generate the corresponding HR residual patches. Let residual_h denote the resulting HR residual face image with the average pixel values in overlapping regions. Together with initial_h, we are able to obtain the final super-resolution face image h_{sr} of l by

$$h_{sr} = \text{initial}_h + \text{residual}_h. \tag{16}$$

3. EXPERIMENTS

To verify the effectiveness of our SCFR-FSR method, we compare the proposed SCFR-FSR with methods CLLR-SR [9], partial least squares (PLS) [14], thresholding locality-constrained representation (TLcR) [15], VCRL-ANE [16], VDSR [17], and ReDegNet [18] on the CAS-PEAL-R1 face database. In addition, we also make use of the Bicubic interpolation algorithm as a comparison baseline. Two metrics, i.e., peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), are utilized to evaluate the performance of FSR. Note that, the higher the values of PSNR and SSIM are, the better the performance is.

3.1. Data Preparation

The CAS-PEAL-R1 database includes 30863 images of 1040 individuals with different variations such as expression, lighting, background, and so on. In our experiments, we select 1040 frontal face images for the performance test. We randomly select 1000 images from 1040 faces for training and the remaining 40 images for testing. The size of HR face images is 96×96 pixels. The LR face images are generated by downsampling the corresponding HR faces with two different factors of 2 and 4. As a result, the size of LR faces is, respectively, 48×48 pixels and 24×24 pixels.

3.2. Parameter Setting

Our proposed SCFR-FSR method has three fractional order parameters, i.e., α_h, α_l, and β, which are separately selected from $\{0.1, 0.2, \cdots, 1\}$. We choose kernel mapping $\psi(\cdot)$ as the Gaussian kernel mapping. Hence, the corresponding kernel function is

$$ker(f_a^{t_a}, f_b^{t_b}) = \exp\left(- \left\| f_a^{t_a} - f_b^{t_b} \right\|^2 / (2\sigma^2)\right)$$

where σ denotes the width parameter set to 0.1. In addition, we empirically set different neighborhood parameters under distinct downsampling factors.

3.3. Comparison of FSR Results

The quantitative FSR results of each method are summarized in Table 1. Taking two test face images as an example, Fig. 1 shows the visual results of each method under different downsampling factors. As can be seen from Table 1, our SCFR-FSR method achieves the best quantitative results with different factors among all the methods, whether the used metric is PSNR or SSIM. The TLcR method performs the second best and comparably to our proposed method. The ReDegNet and Bicubic methods perform worse than other methods. From Fig. 1, it can be seen that our proposed method achieves better visual results, particularly in the $4\times$ case. The Bicubic method achieves the worst visual result among all the methods in the case of $4\times$. These results suggest that our proposed SCFR-FSR method is effective and promising for FSR.

Table 1. Average PSNR and SSIM values of different methods with magnification factors 2 and 4 on the CAS-PEAL-R1 database.

Factor	Metric	CLLR-SR	PLS	TLcR	VCRL-ANE	VDSR	ReDegNet	Bicubic	**Ours**
2×	PSNR	33.06	33.30	35.67	33.61	34.09	31.63	32.25	**38.11**
	SSIM	0.9550	0.9503	0.9665	0.9592	0.9650	0.9414	0.9503	**0.9680**
4×	PSNR	28.40	29.45	30.92	28.59	29.46	26.91	26.90	**33.15**
	SSIM	0.8928	0.8977	0.9112	0.8827	0.8875	0.8633	0.8454	**0.9198**

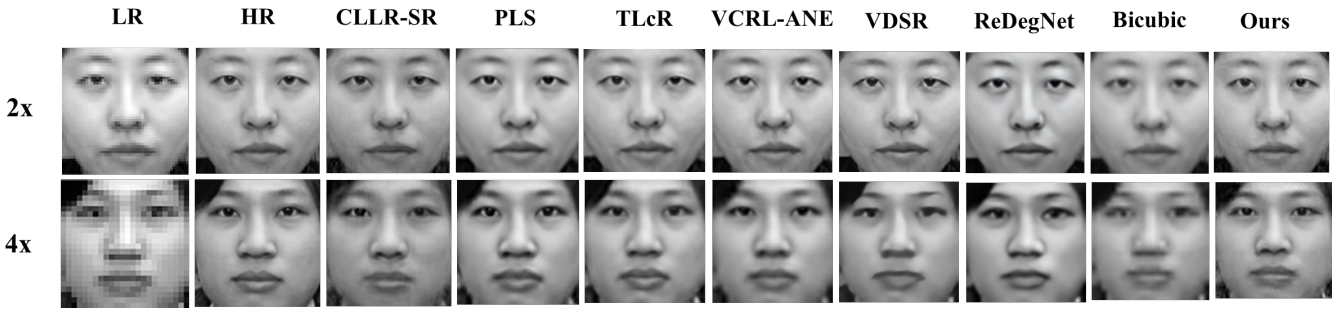

Fig. 1. Visual comparison of face super-resolution results of different methods on the CAS-PEAL-R1 database.

Fig. 2. Visual results of our SCFR-FSR method. (a) LR face images, (b) reconstructed initial HR face images, (c) final SR face images, and (d) original HR face images.

3.4. Further Analysis

We further analyze the performance of the proposed SCFR-FSR using five face images. Fig. 2 shows the reconstructed initial HR face images and final super-resolution face images. As we can see, residual learning strategy improves the facial details of reconstructed initial HR face images and makes the final super-resolution face images closer to original HR faces. This means that it is necessary to introduce residual learning in FSR.

4. CONCLUSION

In this paper, we have proposed a novel SCFR-FSR method for super-resolving face images, where the spectra of sample covariation matrices are reconstructed via using fractional order strategy. Our proposed SCFR-FSR method can not only model the nonlinear relationship between HR and LR facial features, but also alleviates the deviation problem of sample covariation matrices. Experimental results show that the proposed SCFR-FSR method performs better, in contrast with existing methods.

5. REFERENCES

[1] Jinxiu Liang, Jingwen Wang, Yuhui Quan, Tianyi Chen, Jiaying Liu, Haibin Ling, and Yong Xu, "Recurrent exposure generation for low-light face detection," *IEEE Transactions on Multimedia*, vol. 24, pp. 1609–1621, 2022.

[2] Guangwei Gao, Yi Yu, Jian Yang, Guo-Jun Qi, and Meng Yang, "Hierarchical deep cnn feature set-based representation learning for robust cross-resolution face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, pp. 2550–2560, 2022.

[3] Zhiyi Cheng, Xiatian Zhu, and Shaogang Gong, "Face re-identification challenge: Are face recognition models good enough?," *Pattern Recognition*, vol. 107, pp. 107422, 2020.

[4] Jingwen He, Wu Shi, Kai Chen, Lean Fu, and Chao Dong, "Gcfsr: a generative and controllable face super resolution method without facial and gan priors," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2022, pp. 1879–1888.

[5] Xinya Wang, Jiayi Ma, and Junjun Jiang, "Contrastive learning for blind super-resolution via a distortion-specific network," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, pp. 78–89, 2023.

[6] Licheng Liu, C. L. Philip Chen, and Yaonan Wang, "Modal regression-based graph representation for noise robust face hallucination," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, pp. 2490–2502, 2023.

[7] Yuanzhi Wang, Tao Lu, Yanduo Zhang, Zhongyuan Wang, Junjun Jiang, and Zixiang Xiong, "FaceFormer: Aggregating global and local representation for face hallucination," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 6, pp. 2533–2545, 2023.

[8] Harold Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[9] Hua Huang, Huiting He, Xin Fan, and Junping Zhang, "Super-resolution of human face image using canonical correlation analysis," *Pattern Recognition*, vol. 43, no. 7, pp. 2532–2543, 2010.

[10] Le An and Bir Bhanu, "Face image super-resolution using 2d cca," *Signal Processing*, vol. 103, pp. 184–194, 2014.

[11] Yun-Hao Yuan, Jin Li, Yun Li, Jipeng Qiang, Yi Zhu, Xiaobo Shen, and Jianping Gou, "Learning canonical $\mathcal{F}$-correlation projection for compact multiview representation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2022, pp. 19238–19247.

[12] Yun-Hao Yuan, Quan-Sen Sun, and Hong-Wei Ge, "Fractional-order embedding canonical correlation analysis and its applications to multi-view dimensionality reduction and recognition," *Pattern Recognition*, vol. 47, no. 3, pp. 1411–1424, 2014.

[13] Yun-Hao Yuan, Yun Li, Jipeng Qiang, and Bin Li, "Fractional generalized multiview discriminative projections and extension for multiview learning," *Cognitive Systems Research*, vol. 58, pp. 19–30, 2019.

[14] Yuanhong Hao and Chun Qi, "Modified neighbor embedding-based face hallucination using coupled mappings of partial least squares," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 3906–3909.

[15] Junjun Jiang, Yi Yu, Suhua Tang, Jiayi Ma, Akiko Aizawa, and Kiyoharu Aizawa, "Context-patch face hallucination based on thresholding locality-constrained representation and reproducing learning," *IEEE Transactions on Cybernetics*, vol. 50, no. 1, pp. 324–337, 2020.

[16] Guangwei Gao, Yi Yu, Huimin Lu, Jian Yang, and Dong Yue, "Context-patch representation learning with adaptive neighbor embedding for robust face image super-resolution," *IEEE Transactions on Multimedia*, vol. 25, pp. 1879–1889, 2023.

[17] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 1646–1654.

[18] Xiaoming Li, Chaofeng Chen, Xianhui Lin, Wangmeng Zuo, and Lei Zhang, "From face to natural image: Learning real degradation for blind image super-resolution," in *European Conference on Computer Vision (ECCV)*. Springer, 2022, pp. 376–392.