

Special Topic: Logical System Control

Semi-tensor product-based convolutional neural network

Daizhan CHENG¹ & Xiao ZHANG^{2*}¹*Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing 100190, China*²*Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong 999077, China*

Received 13 October 2025/Revised 25 November 2025/Accepted 18 December 2025/Published online 10 March 2026

Abstract The semi-tensor product (STP) of vectors generalizes the conventional inner product, enabling algebraic operations between vectors of different dimensions. Building upon this foundation, we introduce a domain-based convolutional product (CP) and integrate it with the STP to formulate a padding-free convolutional operation. This new operation inherently avoids zero or other artificial padding, thereby eliminating redundant information and boundary artifacts commonly present in conventional convolutional neural networks (CNNs). Based on this operation, we further develop an STP-based CNN framework that extends convolutional computation to irregular and cross-dimensional data domains. Applications to image processing and third-order signal identification demonstrate the proposed method's effectiveness in handling irregular, incomplete, and high-dimensional data without the distortions caused by padding.

Keywords semi-tensor product, domain-based convolution, convolutional neural network, irregular image, higher-order signal identification

Citation Cheng D Z, Zhang X. Semi-tensor product-based convolutional neural network. *Sci China Inf Sci*, 2026, 69(4): 140201, <https://doi.org/10.1007/s11432-025-4734-4>

1 Introduction

Convolutional neural networks (CNNs) have become a cornerstone of modern artificial intelligence, delivering state-of-the-art performance across domains such as computer vision, speech recognition, and natural language understanding. From the early success of gradient-based learning and large-scale image classification to the breakthroughs in residual networks and AlphaGo, convolution-based architectures have consistently demonstrated strong inductive biases for spatial and spatiotemporal data [1–5]. In typical vision pipelines, convolution computes local weighted aggregations over fixed-size receptive fields, enabling hierarchical feature extraction with translation equivariance. However, in finite domains and real-world datasets with boundaries, irregularities, or missing regions, conventional convolution relies on explicit padding (e.g., zero, mirror, or circular), masking, or ad hoc fixes to maintain output dimensionality and kernel alignment. These auxiliary elements are not part of the true observation and may introduce artifacts or statistical bias, thereby degrading feature purity and downstream task performance.

To mitigate boundary effects and handle data incompleteness, a wide range of operator-level enhancements has been developed. Dilated convolutions and multi-scale mechanisms expand receptive fields without increasing parameters [6]; deformable convolutions learn spatial offsets to adapt sampling to content [7]; and sparse or submanifold convolutions improve efficiency on irregular supports [8]. For free-form holes and masks, partial and gated convolutions dynamically reweight valid entries to improve inpainting and reconstruction [9, 10]. Complementary to these algorithmic advances, matrix-based acceleration techniques (e.g., *im2col*) and optimized computational primitives have enhanced deployment efficiency in industrial libraries [11, 12]. Despite these achievements, most approaches remain fundamentally dependent on fixed-dimensional vector or tensor operations. When the number of valid samples within a receptive field varies spatially or according to data quality, the conventional inner product and tensor calculus cannot naturally accommodate cross-dimensional interactions without padding or explicit masking. Beyond practical inefficiency, the reliance on padding also reflects a theoretical inconsistency: convolutional operations are locally linear but globally constrained by fixed tensor shapes. This rigidity prevents convolution from being a truly domain-aware operator: its algebra is defined in Euclidean grid space rather than over the actual support of valid data. Consequently, once the receptive field overlaps incomplete or irregular regions, conventional CNNs lose their algebraic consistency, forcing heuristic treatments such as masking or pseudo-padding.

* Corresponding author (email: xiaozhang@amss.ac.cn)

Motivated by these limitations, this paper introduces a new convolutional framework grounded in the semi-tensor product (STP) of vectors, which enables algebraically consistent operations between vectors of different dimensions [13–15]. While deformable and sparse convolutions relax spatial constraints, they still operate within fixed-dimensional tensor spaces. In contrast, the proposed STP-based convolution operates in \mathbb{R}^∞ , providing a principled, dimension-free formulation that inherently accommodates variable receptive field sizes without the need for explicit padding or masking. From a mathematical viewpoint, the STP provides a natural bridge between continuous-domain algebra and discrete tensor computation. It extends the conventional tensor inner product to a cross-dimensional setting, preserving linearity and associativity while removing the dependency on fixed array dimensions. This property makes it uniquely suitable for modeling convolution over irregular, incomplete, or multi-resolution domains, where the notion of a fixed grid is no longer adequate.

To preserve both theoretical rigor and engineering practicality, we employ a block Hadamard product and a receptive field matrix (RFM) to express the sliding-window computation in a unified linear-algebraic form. This construction yields an STP-based CNN (STP-CNN) whose convolutional operation is defined by the inner product between variable-length valid token vectors and fixed kernels. The resulting framework generalizes naturally to irregular images, partially damaged inputs, and proportional convolution, where receptive field and kernel sizes are mismatched. Furthermore, the method is extended to third-order signals (e.g., 3D medical imaging) via depth-wise masking and rearrangement, enabling effective processing of volumetric and multi-channel data [16–18].

The main contributions of this work are as follows.

- **Padding-free convolution operator.** We propose a convolution operator based on the STP that computes cross-dimensional inner products solely over valid data. This design eliminates zero-padding and explicit mask propagation, thereby reducing boundary artifacts and enhancing robustness to irregular or partially missing inputs.

- **Unified linear-algebraic formulation.** A concise formulation based on the block Hadamard product and RFM expresses the STP convolution in matrix form, maintaining compatibility with mainstream deep learning implementations while grounded in rigorous STP theory.

- **Scalable generalization to 3D and mismatched scales.** The framework naturally supports proportional convolution (where receptive field and kernel sizes differ) and extends to volumetric data via depth-wise masking and rearrangement, unifying 2D irregular images, damaged inputs, and 3D signals within a single paradigm.

In summary, we propose an STP-CNN that replaces padding-dependent inner products with cross-dimensional STP operations and domain-restricted convolution. This approach preserves the mathematical essence of convolution while minimizing the introduction of extraneous information, demonstrating strong adaptability across 2D irregular images, partially damaged inputs, and 3D volumetric signals. We believe this direction offers a rigorous and practical pathway toward representation learning on high-dimensional, irregular, and incomplete data. Future work will focus on large-scale empirical evaluation, efficient integration with mainstream libraries [11, 12], and systematic comparisons with deformable, sparse, and gated convolutional baselines.

The remainder of this paper is organized as follows. Section 2 reviews the STP and the cross-dimensional inner product on \mathbb{R}^∞ as the algebraic foundation. Section 3 defines the domain-based convolutional product (DCP) and establishes its properties. Section 4 introduces the block Hadamard product and RFM. Section 5 formulates the STP-based convolutional product (STP-CP) operating without padding. Section 6 demonstrates STP-CP on irregular and partially damaged images as well as proportional convolution. Section 7 extends the framework to 3D signals with depth masking and rearrangement. Section 8 concludes and outlines directions for large-scale experimentation and library integration.

2 Semi-tensor product of vectors

To establish the algebraic foundation of the proposed framework, this section reviews the STP of vectors and the associated cross-dimensional inner product defined on \mathbb{R}^∞ . The STP enables consistent operations between vectors of different dimensions, overcoming the fixed-size limitation of conventional inner products. We further introduce the corresponding norm, distance, and equivalence relations, which together form a topological vector space structure that will serve as the mathematical basis for subsequent developments.

Consider the cross-dimensional Euclidean space

$$\mathbb{R}^\infty := \bigcup_{n=1}^{\infty} \mathbb{R}^n.$$

Definition 1. Let $x \in \mathbb{R}^m \subset \mathbb{R}^\infty$ and $y \in \mathbb{R}^n \subset \mathbb{R}^\infty$, with $t = \text{lcm}(m, n)$. The addition (or subtraction) of x and y is defined as

$$x \overset{\rightarrow}{\pm} y := (x \otimes \mathbf{1}_{t/m}) \pm (y \otimes \mathbf{1}_{t/n}) \in \mathbb{R}^t, \tag{1}$$

where \otimes denotes the Kronecker product and $\mathbf{1}_k$ is a column vector of ones of length k .

Definition 2. Let $x \in \mathbb{R}^m \subset \mathbb{R}^\infty$ and $y \in \mathbb{R}^n \subset \mathbb{R}^\infty$.

(i) **Inner product:**

$$\langle x, y \rangle_{\mathcal{V}} := \frac{1}{t} \langle (x \otimes \mathbf{1}_{t/m}), (y \otimes \mathbf{1}_{t/n}) \rangle, \tag{2}$$

where $t = \text{lcm}(m, n)$ and $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product.

(ii) **Norm:**

$$\|x\|_{\mathcal{V}} := \sqrt{\langle x, x \rangle_{\mathcal{V}}}. \tag{3}$$

(iii) **Distance:**

$$d_{\mathcal{V}}(x, y) := \|x \overset{\rightarrow}{-} y\|_{\mathcal{V}}. \tag{4}$$

Lemma 1 ([13]). (i) Under the addition defined in (1) and the conventional scalar multiplication, \mathbb{R}^∞ forms a pseudo-vector space. (ii) Equipped with the topology induced by the distance in (4), \mathbb{R}^∞ is a topological space.

Definition 3. For $x, y \in \mathbb{R}^\infty$, we say x and y are equivalent, denoted by $x \leftrightarrow y$, if

$$d_{\mathcal{V}}(x, y) = 0.$$

The equivalence class of x is then defined as

$$\bar{x} := \{y \mid y \leftrightarrow x\}. \tag{5}$$

Definition 4. Let Ω denote the set of equivalence classes

$$\Omega = \mathbb{R}^\infty / \leftrightarrow := \{\bar{x} \mid x \in \mathbb{R}^\infty\}.$$

For $\bar{x}, \bar{y} \in \Omega$, we define the following.

(i) **Addition:**

$$\bar{x} \overset{\rightarrow}{\pm} \bar{y} := x \overset{\rightarrow}{\pm} y, \quad \bar{x}, \bar{y} \in \Omega. \tag{6}$$

(ii) **Inner product:**

$$\langle \bar{x}, \bar{y} \rangle_{\mathcal{V}} := \langle x, y \rangle_{\mathcal{V}}, \quad \bar{x}, \bar{y} \in \Omega. \tag{7}$$

(iii) **Norm:**

$$\|\bar{x}\|_{\mathcal{V}} := \|x\|_{\mathcal{V}}. \tag{8}$$

(iv) **Distance:**

$$d_{\mathcal{V}}(\bar{x}, \bar{y}) := d_{\mathcal{V}}(x, y). \tag{9}$$

Lemma 2 ([15]). (i) The definitions in (6)–(9) are well-defined, i.e., they are independent of the choice of representative elements. (ii) Under these operations and the topology induced by the distance in (9), Ω forms a topological vector space (see [19] for details).

Figure 1 illustrates the geometric structure of Ω in comparison to the conventional Euclidean structure, where \mathcal{T}_n denotes the natural topology and \mathcal{T}_d represents the topology induced by the distance.

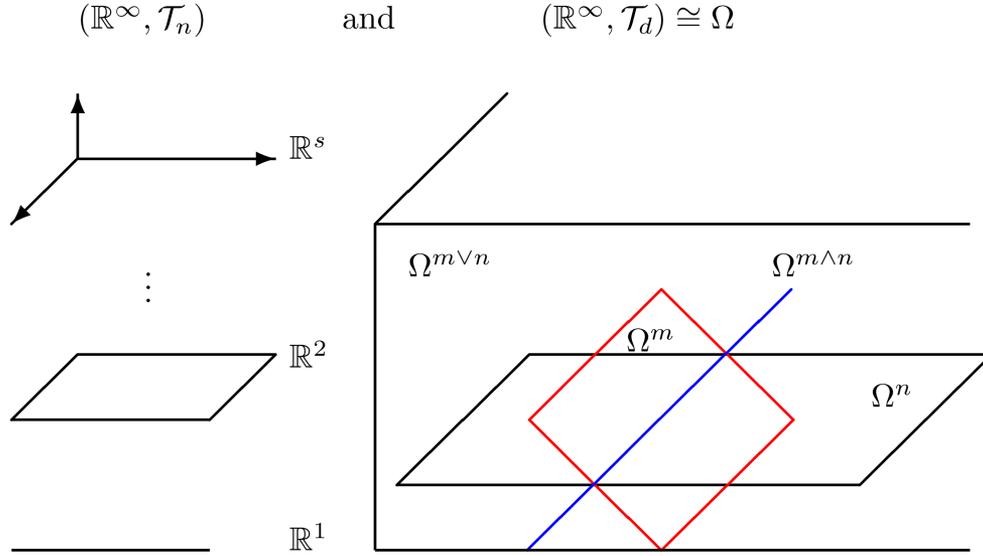


Figure 1 (Color online) $(\mathbb{R}^\infty, \mathcal{T}_n)$ vs. Ω .

3 Classical CP and domain-based CP

Building on the preliminaries of the STP, this section revisits the classical convolutional product (CP) in both continuous and discrete forms and extends it to DCP. Unlike the traditional CP, DCP restricts the integration or summation to the valid data region, which naturally aligns with real-world cases where signals or images contain missing or irregular portions. This section also summarizes several key properties of DCP, preparing the groundwork for its matrix-form implementation in Section 4.

3.1 Classical CP

Definition 5. Let $f(x), g(x) \in L^1(\mathbb{R})$. The CP of $f(x)$ and $g(x)$, denoted by $f(x) * g(x)$, is defined as

$$s(x) := f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau) g(x - \tau) d\tau. \tag{10}$$

Some fundamental properties of CP are summarized below.

Proposition 1 ([20]). Key properties of the CP include the following.

(i) L^1 -Stability:

$$\|f(x) * g(x)\|_1 \leq \|f(x)\|_1 \|g(x)\|_1. \tag{11}$$

Hence, $f(x) * g(x) \in L^1(\mathbb{R})$.

(ii) Commutativity:

$$f(x) * g(x) = g(x) * f(x). \tag{12}$$

(iii) Associativity: if $f(x), g(x), h(x) \in L^1(\mathbb{R})$, then

$$f(x) * (g(x) * h(x)) = (f(x) * g(x)) * h(x). \tag{13}$$

The above properties concern the continuous setting. For discrete-time signals, convolution is defined as follows.

Definition 6. Let $f(n), w(n) \in \ell^1$. The discrete-time CP of $f(n)$ and $w(n)$ is

$$s(n) := f(n) * w(n) = \sum_{\tau=-\infty}^{\infty} f(\tau) w(n - \tau). \tag{14}$$

When w is regarded as a weighting function, Eqs. (10) and (14) can be interpreted as a weighted average, where the weights depend on the distance between the current position and the arguments of the function being averaged. Proposition 1 also holds for the discrete-time CP. Using (11), we further obtain the following result.

Proposition 2. The stability results of Proposition 1 extend directly to the discrete-time case and finite windows:

- (i) If $f(x), w(x) \in L^1$, then $f(x) * w(x) \in L^1$;
- (ii) If $f(n), w(n) \in \ell^1$, then $f(n) * w(n) \in \ell^1$.

3.2 CP with finite windows

In deep learning and many signal processing tasks, convolution is computed over a finite window:

$$s(n) := f(n) * k(n) = \sum_{\tau=-N}^N f(\tau) k(n - \tau), \tag{15}$$

where $[-N, N]$ is the sampling window, $f(n)$ is the input signal, and $k(n)$ is the convolution kernel.

By exploiting the commutativity property in (12), one may alternatively write

$$s(n) := f(n) * k(n) = \sum_{\tau=-N}^N f(n - \tau) k(\tau). \tag{16}$$

In CNNs, a closely related operation—cross-correlation—is often used, typically defined as

$$s(n) := f(n) * k(n) = \sum_{\tau=-N}^N f(n + \tau) k(\tau). \tag{17}$$

Note. although the same symbol “*” is often used in practice, Eq. (17) corresponds to cross-correlation rather than the commutative convolution in (15) and (16).

For multivariate functions, CP generalizes naturally. For two-variable functions,

$$s(m, n) := f(m, n) * k(m, n) = \sum_i \sum_j f(i, j) k(m - i, n - j), \tag{18}$$

or equivalently,

$$s(m, n) := f(m, n) * k(m, n) = \sum_i \sum_j f(m - i, n - j) k(i, j). \tag{19}$$

This two-dimensional CP is standard in image processing.

3.3 Domain-based CP

Definition 7. Let $D \subset \mathbb{R}$ be measurable, $f(x) \in L^1(D)$, and $w(x) \in L^1(\mathbb{R})$. The domain-based convolutional product (DCP) of $f(x)$ with $w(x)$ is

$$s(x) := f(x) * w(x) = \int_{\tau \in D} f(\tau) w(x - \tau) d\tau. \tag{20}$$

Several basic properties of classical CP extend directly to DCP.

Proposition 3. (i) The domain-based convolution $f(x) * w(x)$ is a measurable function on \mathbb{R} . Moreover, we have

$$\|f(x) * w(x)\|_1 \leq \|f(x)\|_1^D \|w(x)\|_1 < \infty. \tag{21}$$

Therefore, $f(x) * w(x) \in L^1(\mathbb{R})$.

(ii) The commutativity relation holds in the following sense:

$$\int_{\tau \in D} f(\tau) w(t - \tau) d\tau = \int_{t - \tau \in D} w(\tau) f(t - \tau) d\tau. \tag{22}$$

(iii) If $f(x) \in L^1(D)$ and $g(x), h(x) \in L^1(\mathbb{R})$, then

$$f(x) * (g(x) * h(x)) = (f(x) * g(x)) * h(x). \tag{23}$$

Remark 1. The proofs of Proposition 3 follow straightforwardly from Definition 7 and the properties of the classical convolution. They are therefore omitted.

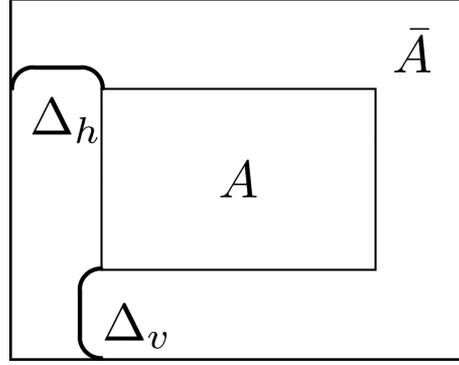


Figure 2 Illustration of image padding to obtain \bar{A} .

4 CP for image processing

This section develops a matrix-based representation of classical CNN convolution, establishing a unified algebraic framework for the subsequent derivation. We first recall the standard procedure of image convolution with zero-padding and then introduce two key matrix constructs: the block Hadamard product and the RFM. In classical CNNs, the receptive field is an implicit, sliding-window concept referring to the local input region (e.g., an $s \times t$ patch) processed by the kernel. The RFM R_A provides an explicit matrix representation by gathering all receptive fields of the padded image \bar{A} into a structured block matrix. These tools not only clarify the linear-algebraic essence of convolution but also pave the way for its extension to STP-based operations in Section 5.

4.1 Classical CP in CNNs for image processing

The CP used in CNNs for image processing is well documented in standard references (see, e.g., [4, 21, 22]).

Given an image matrix $A \in \mathbb{R}^{m \times n}$ and a kernel matrix $K \in \mathbb{R}^{s \times t}$, the convolution proceeds as follows.

Algorithm 1. Step 1: padding. To address boundary effects, A is padded (typically with zeros) so that all original pixels become interior points. Let the padded matrix be $\bar{A} \in \mathbb{R}^{p \times q}$, where

$$p = m + 2\Delta_v, \quad q = n + 2\Delta_h, \quad (24)$$

and

$$\bar{a}_{i,j} = \begin{cases} a_{i-\Delta_v, j-\Delta_h}, & \Delta_v + 1 \leq i \leq \Delta_v + m, \\ & \Delta_h + 1 \leq j \leq \Delta_h + n, \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

This process is illustrated in Figure 2.

Step 2: local convolution. For each receptive field $\bar{A}_{i,j}$ of \bar{A} covered by the kernel K , compute

$$s_{i,j} = \langle V_c(K), V_c(\bar{A}_{i,j}) \rangle, \quad (26)$$

where $V_c(\cdot)$ denotes the column stacking form of a matrix.

Step 3: sliding and output construction. Slide K across \bar{A} with vertical and horizontal strides d_v and d_h , respectively, to form $S = (s_{i,j}) \in \mathbb{R}^{s_v \times s_h}$ satisfying

$$\begin{cases} (s_v - 1)d_v + s = p, \\ (s_h - 1)d_h + t = q. \end{cases} \quad (27)$$

The padding sizes Δ_v and Δ_h are chosen so that Eq. (27) holds.

The following example demonstrates the algorithm with a concrete case.

Example 1. Consider a grayscale image $A \in \mathbb{R}^{3 \times 4}$ and a kernel $K \in \mathbb{R}^{2 \times 2}$, where

$$A = \begin{bmatrix} 1 & 2 & -1 & -2 \\ -3 & -2 & 1 & 3 \\ 2 & -2 & 1 & -1 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 0.4 \\ 0.6 & 1.5 \end{bmatrix}.$$

With $\Delta_v = \Delta_h = 1$, the padded image is $\bar{A} \in \mathbb{R}^{5 \times 6}$. Using (26), each $s_{i,j}$ is computed by

$$s_{s,t} = \sum_{i=1}^2 \sum_{j=1}^2 \bar{a}_{s-2+i,t-2+j} k_{i,j}, \quad 2 \leq s \leq 4, \quad 2 \leq t \leq 5. \quad (28)$$

The resulting CP matrix is

$$S = \begin{bmatrix} 1.5 & 3.6 & -0.3 & -3.6 & -1.2 \\ -4.1 & -3.0 & 1.9 & 3.2 & -0.2 \\ 1.8 & -5.6 & -1.3 & 1.3 & 2.4 \\ 0.8 & 1.2 & -1.6 & 0.6 & -1 \end{bmatrix}.$$

Figure 3 shows a schematic of the input, kernel, and selected outputs (e.g., $s_{1,1}, s_{2,2}, s_{4,4}$).

4.2 Block Hadamard product and receptive field matrix

To compute CP efficiently, we introduce two matrix structures: the block Hadamard product and the RFM.

Definition 8 (Block Hadamard product). Let A and B be two matrices of identical size and block structure:

$$A = \begin{bmatrix} A^{1,1} & \cdots & A^{1,q} \\ \vdots & \ddots & \vdots \\ A^{p,1} & \cdots & A^{p,q} \end{bmatrix}, \quad B = \begin{bmatrix} B^{1,1} & \cdots & B^{1,q} \\ \vdots & \ddots & \vdots \\ B^{p,1} & \cdots & B^{p,q} \end{bmatrix}. \quad (29)$$

The block Hadamard product of A and B , denoted by $A \bar{\circ} B$, is defined as

$$A \bar{\circ} B := C = (c_{i,j}) \in \mathbb{R}^{p \times q}, \quad (30)$$

where

$$c_{i,j} = \langle V_c(A^{i,j}), V_c(B^{i,j}) \rangle, \quad i \in [1, p], \quad j \in [1, q].$$

Definition 9 (RFM). Let $\bar{A} \in \mathbb{R}^{p \times q}$ be a padded image satisfying (27). Define the RFM R_A as

$$R_A = P \bar{A} Q \in \mathbb{R}^{(s_v s) \times (s_h t)}, \quad (31)$$

where

$$P^T = \Xi_{s_v \times s}^{d_v}, \quad Q = \Xi_{s_h \times t}^{d_h}. \quad (32)$$

Here, $\Xi_{n \times \eta}^d \in \mathbb{R}^{\xi \times (n\eta)}$ is a structured selection matrix with stride d , $\xi = (n-1)d + \eta$, as illustrated in Figure 4.

The matrix R_A admits a block form

$$R_A = \begin{bmatrix} R_{1,1} & \cdots & R_{1,s_h} \\ \vdots & \ddots & \vdots \\ R_{s_v,1} & \cdots & R_{s_v,s_h} \end{bmatrix}, \quad (33)$$

where each $R_{i,j} \in \mathbb{R}^{s \times t}$ corresponds to a receptive field of \bar{A} . Then the CP matrix entries are

$$s_{i,j} = \langle V_c(K), V_c(R_{i,j}) \rangle, \quad i \in [1, s_v], \quad j \in [1, s_h]. \quad (34)$$

Using the block Hadamard product, the entire convolutional output can be compactly expressed as follows.

Proposition 4. Let R_A be defined as in (33), where each block $R_{i,j}$ matches the size of K . Then the convolutional output matrix is

$$S = (\mathbf{1}_{s_v \times s_h} \otimes K) \bar{\circ} R_A. \quad (35)$$

Proof. This follows directly from the construction of the RFM R_A in (33) and the definition of the block Hadamard product in (30), which together implement the sliding-window inner product described in Algorithm 1.

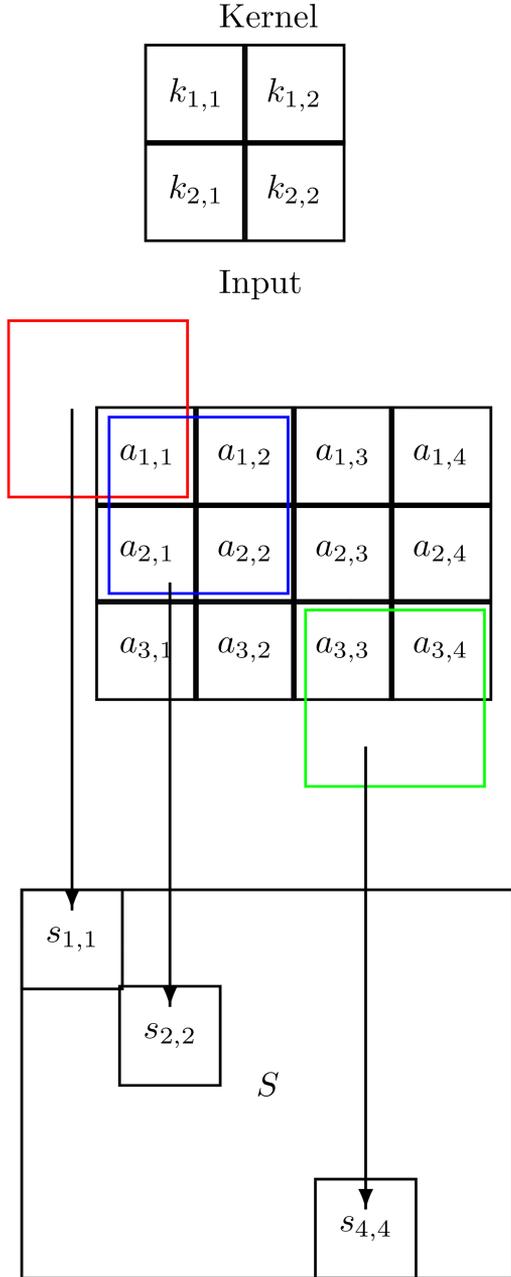


Figure 3 (Color online) Image CP.

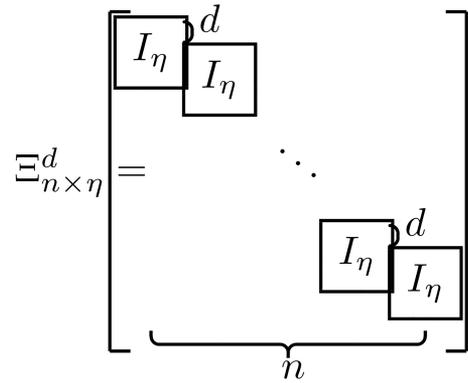


Figure 4 Matrix $\Xi_{n \times \eta}^d$.

5 Semi-tensor product-based CP

With the mathematical and structural groundwork in place, this section formally defines the STP-CP. The proposed operation replaces the conventional inner product with the STP-based cross-dimensional inner product, allowing convolution to be performed directly on valid entries without zero-padding or explicit mask propagation. Through examples and remarks, we illustrate how STP-CP generalizes convolution to variable-sized receptive fields, enabling a padding-free and dimension-flexible CNN formulation.

We recall the STP of vectors defined in (2) for completeness. In CP of CNNs, each element of the CP matrix S is computed via a conventional inner product, as in (26).

In the proposed STP-CP, consider a receptive field $\bar{A}_{i,j}$. If certain entries in this receptive field are undefined (e.g., at image boundaries or within damaged regions), we do not pad them with zeros; instead, we simply ignore the invalid entries. Consequently, different receptive fields may have different effective dimensions, and the STP of

vectors of varying lengths is used to compute the inner product between K and $A_{i,j}$. This yields the STP-CNN, a convolutional neural network constructed using the STP-CP. Formally, it is defined as follows.

Definition 10. Let A be the image matrix and K the kernel matrix. The STP-based convolutional product is

$$s_{i,j} := \langle V_c(A_{i,j}), V_c(K) \rangle_{\mathcal{V}}, \quad (36)$$

where $A_{i,j}$ denotes the subset of valid (available) entries within the receptive field corresponding to position (i, j) .

Remark 2. Referring to Figure 3, consider several receptive fields of varying dimensions: the field for $s_{1,1}$ is a 1×1 matrix $A_{1,1} \in \mathbb{R}^{1 \times 1}$, that for $s_{2,2}$ is $A_{2,2} \in \mathbb{R}^{2 \times 2}$, and for $s_{4,4}$ it is $A_{4,4} \in \mathbb{R}^{4 \times 4}$. Because these receptive fields have different dimensions, their corresponding inner products with K are computed via the STP of vectors of different sizes.

Example 2. Continuing Example 1, applying the STP-CP as in (36) yields

$$S = \frac{1}{4} \begin{bmatrix} 3.5 & 5.4 & 1.3 & -5.4 & -7.0 \\ -4.1 & -3 & 1.9 & 3.3 & 2.5 \\ -1 & -5.6 & -1.3 & 1.3 & 2.9 \\ 7.0 & -0.4 & -0.7 & -0.7 & -3.5 \end{bmatrix}.$$

Remark 3 (Computational complexity of STP-CP). To provide a formal comparison, we analyze the computational profile of the proposed STP-CP operation, which generalizes classical convolution. Classical convolution corresponds to the special case of STP-CP where all receptive fields are normal (i.e., of the full $s \times t$ size).

The complexity of classical CP for a single output element with an $s \times t$ kernel is $O(C_{\text{in}} \cdot s \cdot t)$. The overall complexity for an entire layer is $O(O_h \cdot O_w \cdot C_{\text{in}} \cdot C_{\text{out}} \cdot s \cdot t)$, where $O_h \times O_w$ is the output size, C_{in} is the number of input channels, and C_{out} is the number of output channels (i.e., filters).

For STP-CP, we consider two scenarios regarding the receptive fields (the local input regions processed by the kernel).

- **Normal (full-sized) receptive fields.** When the receptive field is complete, STP-CP is computationally identical to classical CP for a single output element, i.e., $O(C_{\text{in}} \cdot s \cdot t)$.

- **Abnormal (irregular) receptive fields.** For a receptive field with v valid entries ($v < s \cdot t$), STP-CP incurs an overhead, with a per-element complexity of $O(C_{\text{in}} \cdot \ell)$, where $\ell = \text{lcm}(v, s \cdot t)$. This overhead stems from calculating ℓ and performing the requisite Kronecker products.

Crucially, since the number of abnormal receptive fields is typically small (e.g., confined to boundaries or sparse damaged regions), the overall additional cost is amortized over the entire input and does not alter the asymptotic complexity of the layer. Furthermore, the parameter count of an STP-CNN layer remains exactly the same as its standard CNN counterpart.

6 Application of STP-CP

This section demonstrates the practical capabilities of the proposed STP-CP framework through several representative examples. We first show how STP-CP operates effectively on irregular and partly damaged images by directly using available data, followed by an application to proportional STP-CP, which handles mismatched receptive field and kernel sizes. These examples collectively verify that STP-CP not only eliminates artifacts introduced by padding but also improves robustness to data irregularity and scale variation.

6.1 Irregular image

The STP-CP framework naturally handles irregular images (those with undefined or missing pixels) by operating directly on valid entries without padding or manual repair.

Example 3. Consider the irregular image in Figure 5, with receptive fields indicated by blue rectangles.

Assume the convolution kernel K is the same as in Figure 3, and let the vertical and horizontal paddings and strides satisfy

$$\Delta_v = \Delta_h = d_v = d_h = 1.$$

Then,

$$P^T = Q = \Xi_{5 \times 2}^1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Using (31), the extended image matrix is represented as

$$M_{\bar{A}} = P\bar{A}Q := (\bar{A}^{i,j}) \in \mathbb{R}^{10 \times 10},$$

where

$$\bar{A}^{i,j} \in \mathbb{R}^{2 \times 2}, \quad i, j \in [1, 5].$$

Let $V_{i,j}$ denote the valid (non-missing) entries of $\bar{A}^{i,j}$. The matrix of valid regions is

$$V = (V_{i,j}) = \begin{bmatrix} \times & 1 & \begin{bmatrix} 1 \\ -1 \end{bmatrix} & -1 & \times \\ -2 & \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ -1 \\ 2 \end{bmatrix} & \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} & 1 \\ \begin{bmatrix} -2 \\ -3 \end{bmatrix} & \begin{bmatrix} -2 \\ -3 \\ 1 \\ 2 \end{bmatrix} & \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix} & \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} & 1 \\ \begin{bmatrix} -3 \\ 2 \end{bmatrix} & \begin{bmatrix} -3 \\ 2 \\ 2 \\ -2 \end{bmatrix} & \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix} & 3 & \times \\ 2 & \begin{bmatrix} 2 \\ -2 \end{bmatrix} & -2 & \times & \times \end{bmatrix},$$

where “ \times ” denotes undefined entries.

Applying the STP-CP to these irregular receptive fields yields

$$\begin{aligned} S &= (\mathbf{1}_{5 \times 5} \otimes K) \bar{\sigma}V \\ &= \frac{1}{12} \begin{bmatrix} \times & 10.5 & -0.9 & -10.5 & \times \\ -21.0 & -0.3 & 12.6 & 5.3 & 10.5 \\ -26.7 & -1.2 & 22.5 & 18.1 & 10.5 \\ -3.0 & -12.0 & 17.9 & 31.5 & \times \\ 21.0 & -1.8 & -21.0 & \times & \times \end{bmatrix}. \end{aligned} \tag{37}$$

This example demonstrates that STP-CP can perform convolution directly on irregular or incomplete image data without the need for zero-padding. By dynamically adjusting the inner product dimensionality through the STP of vectors, the proposed method naturally adapts to the valid regions of the image, preserving structural information and avoiding artificial noise introduced by padding.

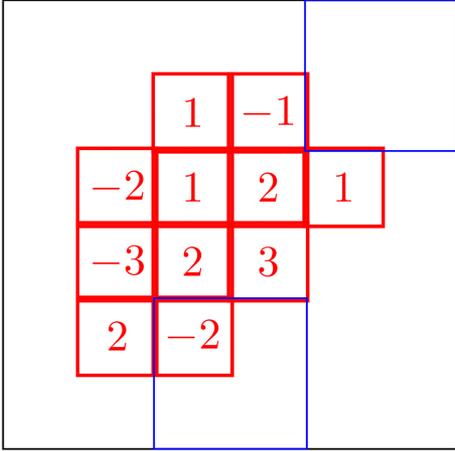


Figure 5 (Color online) An irregular image and its receptive fields.

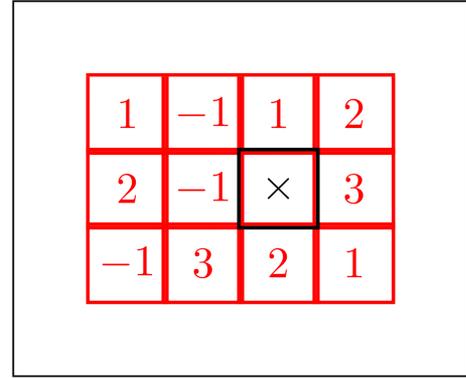


Figure 6 (Color online) STP-CP applied to a partly damaged image.

6.2 Partly damaged image

STP-CP also applies to partly damaged images by restricting computation to available pixels within each receptive field.

Example 4. Consider the partly damaged image in Figure 6, where the damaged region is indicated by “×”.

Assume the kernel K is the same as in Figure 3. Omitting intermediate steps, the vector forms of all receptive fields are

$$V = (V_{i,j}) = \begin{bmatrix} 1 & \begin{bmatrix} 1 \\ -1 \end{bmatrix} & \begin{bmatrix} -1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 2 \end{bmatrix} & 2 \\ \begin{bmatrix} 1 \\ 2 \end{bmatrix} & \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \end{bmatrix} & \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} & \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} & \begin{bmatrix} 2 \\ 3 \end{bmatrix} \\ \begin{bmatrix} 2 \\ -1 \end{bmatrix} & \begin{bmatrix} 2 \\ -1 \\ -1 \\ 3 \end{bmatrix} & \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} & \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} & \begin{bmatrix} 3 \\ 1 \end{bmatrix} \\ -1 & \begin{bmatrix} -1 \\ 3 \end{bmatrix} & \begin{bmatrix} 3 \\ 2 \end{bmatrix} & \begin{bmatrix} 2 \\ 1 \end{bmatrix} & 1 \end{bmatrix}.$$

Applying STP-CP to the above, the resulting CP matrix is

$$S = \frac{1}{12} \begin{bmatrix} 10.5 & -0.9 & 0.9 & 16.2 & 21 \\ 16.2 & 0.9 & -0.7 & 22.3 & 26.7 \\ 3.9 & 16.5 & 12.2 & 18.1 & 20.1 \\ -10.5 & 12.3 & 25.8 & 15.3 & 10.5 \end{bmatrix}. \tag{38}$$

6.3 Proportional STP-CP

As mentioned in [22], “Assume the size of a dog is the same as the training dogs, CNN can recognize it well. But if the dog image is considerably enlarged, CNN may fail to recognize it.” Conventional CNNs often struggle with large-scale variations: a model trained on objects of one size may fail when objects are significantly larger or smaller. STP-CP addresses this by allowing receptive fields whose sizes do not match the kernel size exactly. By setting receptive field sizes appropriately, a proportional STP-CNN can recognize objects across scales. The following example illustrates proportional STP-CP for mismatched image sizes.

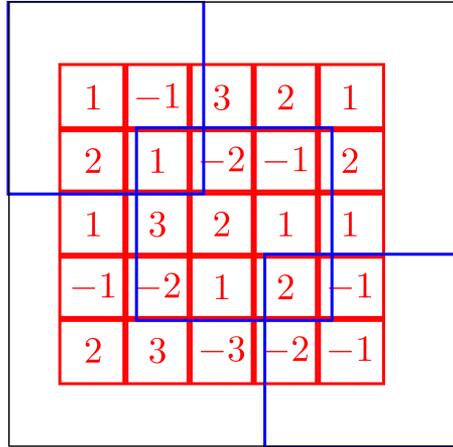


Figure 7 (Color online) Proportional STP-CP for mismatched sizes.

Example 5. Consider the enlarged image shown in Figure 7. Use a receptive field of size 3×3 and a step length of 2; diagonal receptive fields are highlighted in blue.

With the same kernel K as in Figure 3, the vector forms of all receptive fields are

$$V = (V_{i,j}) = \begin{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \\ 1 \end{bmatrix} & \begin{bmatrix} -1 \\ 1 \\ 3 \\ -2 \\ 2 \\ -1 \end{bmatrix} & \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \end{bmatrix} \\ \begin{bmatrix} 2 \\ 1 \\ -1 \\ 1 \\ 3 \\ -2 \end{bmatrix} & \begin{bmatrix} 1 \\ 3 \\ -2 \\ -2 \\ 2 \\ 1 \\ -1 \\ 1 \\ 2 \end{bmatrix} & \begin{bmatrix} -1 \\ 1 \\ 2 \\ 2 \\ 1 \\ -1 \end{bmatrix} \\ \begin{bmatrix} -1 \\ 2 \\ -2 \\ 3 \end{bmatrix} & \begin{bmatrix} -2 \\ 3 \\ 1 \\ -3 \\ 2 \\ -2 \end{bmatrix} & \begin{bmatrix} 2 \\ -2 \\ -1 \\ -1 \end{bmatrix} \end{bmatrix} .$$

Applying STP-CP yields

$$S = \frac{1}{36} \begin{bmatrix} 29.7 & 7.2 & 43.2 \\ 14.7 & 26.5 & -9.9 \\ 35.1 & -7.8 & -7.8 \end{bmatrix} . \tag{39}$$

Remark 4. (i) A grayscale image has a single channel; a color image has three channels (R, G, B) .

(ii) Typically, each convolutional layer contains multiple kernels (filters). With k filters, the number of output channels is k times the number of input channels.

(iii) When the receptive fields are several times larger than the kernel, STP-CP performs a weighted aggregation over an extended region, achieving multi-scale feature extraction within a single layer while remaining padding-free. This property demonstrates the intrinsic scale adaptivity of STP-CP.

7 Cubic image

This section extends the two-dimensional STP-CP framework to three-dimensional data, enabling the processing of cubic images (third-order tensors) from domains including computed tomography (CT), hyperspectral imaging, and spatiotemporal analysis.

A 3D image is a hypermatrix (or tensor, see [18]) A with entries

$$A = \{a_{i,j,k} \mid i \in [1, m], j \in [1, n], k \in [1, \eta]\}. \quad (40)$$

We denote the space of such hypermatrices by $\mathbb{R}^{m \times n \times \eta}$.

Following common practice (see [17, 18]), a cubic image $A \in \mathbb{R}^{m \times n \times \eta}$ can be unfolded into a matrix by stacking the η frontal slices row-wise. Specifically,

$$A \in \mathbb{R}^{m \times n \times \eta} \sim \mathbb{R}^{(\eta m) \times n}, \quad (41)$$

where rows are indexed by (k, i) and columns by j .

We use a numerical example to demonstrate the design process step by step.

Example 6. (1) Cubic image. Consider a cubic image with $m = 3$, $n = 4$, $\eta = 2$ as follows:

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 3 & 2 \\ 1 & 3 & 2 & 2 \\ 3 & 2 & 0 & 1 \\ 1 & 1 & 2 & 3 \\ 4 & 2 & 3 & 4 \\ 4 & 0 & 3 & 3 \end{bmatrix} \in \mathbb{R}^{6 \times 4}. \quad (42)$$

(2) Kernel. A 3D kernel is likewise a cubic tensor, which is represented in matrix form as

$$K \in \mathbb{R}^{s \times t \times \xi} \sim \mathbb{R}^{(\xi s) \times t}. \quad (43)$$

Let $K \in \mathbb{R}^{2 \times 2 \times 3}$ be

$$K = \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & -1 \\ 2 & 3 \\ 2 & 1 \\ 3 & 3 \end{bmatrix} \in \mathbb{R}^{6 \times 2}. \quad (44)$$

(3) Enlarged image. Let the enlarged (padded) image be $\bar{A} \in \mathbb{R}^{p \times q \times \eta}$, where

$$p = m + \Delta m, \quad q = n + \Delta n, \quad (45)$$

where η is unchanged and Δm , Δn are the total increments in the vertical and horizontal sizes, respectively.

Example 7. For Example 6, let $\Delta m = \Delta n = 2$. Then $p = 5$, $q = 6$, $\eta = 2$, and

$$\bar{A} = \begin{bmatrix} \bar{A}_1 \\ \bar{A}_2 \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ \times & 2 & 1 & 3 & 2 & \times \\ \times & 1 & 3 & 2 & 2 & \times \\ \times & 3 & 2 & 0 & 1 & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & 1 & 1 & 2 & 3 & \times \\ \times & 4 & 2 & 3 & 4 & \times \\ \times & 4 & 0 & 3 & 3 & \times \\ \times & \times & \times & \times & \times & \times \end{bmatrix} \in \mathbb{R}^{10 \times 6}. \tag{46}$$

(4) RFM in the spatial plane. Let $n_v \times n_h$ be the vertical and horizontal numbers of receptive fields; let d_v and d_h be the vertical and horizontal strides, and let $s \times t$ be the size of the RFM. Define

$$R = P\bar{A}Q, \tag{47}$$

where

$$\begin{aligned} R &= P\bar{A}Q \in \mathbb{R}^{(n_v s) \times (n_h t)}, \\ Q &= \Xi_{n_h \times t}^{d_h} \in \mathbb{R}^{q \times (n_h t)}, \\ P^T &= \Xi_{n_v \times s}^{d_v} \in \mathbb{R}^{p \times (n_v s)}, \\ (n_h - 1)d_h + t &= q, \\ (n_v - 1)d_v + s &= p. \end{aligned} \tag{48}$$

Example 8. With Examples 6 and 7 and $d_h = d_v = 1$,

$$P^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{49}$$

Then

$$R := \begin{bmatrix} R_1 \\ R_2 \end{bmatrix},$$

where

$$R_i = P\bar{A}_iQ, \quad i = 1, 2.$$

A direct computation gives

$$R = \begin{bmatrix} \times & \times \\ \times & 2 & 2 & 1 & 1 & 3 & 3 & 2 & 2 & \times \\ \times & 2 & 2 & 1 & 1 & 3 & 3 & 2 & 2 & \times \\ \times & 1 & 1 & 3 & 3 & 2 & 2 & 2 & 2 & \times \\ \times & 1 & 1 & 3 & 3 & 2 & 2 & 2 & 2 & \times \\ \times & 3 & 3 & 2 & 2 & 0 & 0 & 1 & 1 & \times \\ \times & 3 & 3 & 2 & 2 & 0 & 0 & 1 & 1 & \times \\ \times & \times \\ \times & \times \\ \times & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 3 & \times \\ \times & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 3 & \times \\ \times & 4 & 4 & 2 & 2 & 3 & 3 & 4 & 4 & \times \\ \times & 4 & 4 & 2 & 2 & 3 & 3 & 4 & 4 & \times \\ \times & 4 & 4 & 0 & 0 & 3 & 3 & 3 & 3 & \times \\ \times & 4 & 4 & 0 & 0 & 3 & 3 & 3 & 3 & \times \\ \times & \times \end{bmatrix} \in \mathbb{R}^{(n_z n_v s) \times (n_h t)}, \quad (50)$$

where n_z is the block number in depth direction (refer to the following (52) for details).

(5) Masking along the depth direction. Let

$$R_0 = \times_{(n_v s) \times (n_h t)},$$

and define

$$\bar{R} = \begin{bmatrix} R_0 \\ R \\ R_0 \end{bmatrix}. \quad (51)$$

Let $\Delta\eta = 2$. Then the masked 3D image should be

$$\bar{R} \in \mathbb{R}^{p \times q \times \delta},$$

where

$$\delta = \eta + \Delta\eta = 4,$$

Moreover, we set

$$R_0 := \times_{8 \times 10},$$

and obtain

$$\bar{R} = \begin{bmatrix} R_0 \\ R_1 \\ R_2 \\ R_0 \end{bmatrix}. \quad (52)$$

(6) RFM along the depth direction. Let ξ be the kernel depth in (43), stride d_z , and n_z the number of depth receptive fields, satisfying

$$(n_z - 1)d_z + \xi = \delta. \quad (53)$$

(8) CP matrix. Finally, partition Ψ into $(n_v n_z) \times n_h$ blocks, each of size $(s\xi) \times t$. Then each block

$$\Psi_{i,j} \in \mathbb{R}^{(s\xi) \times t}, \quad i \in [1, n_v n_z], j \in [1, n_h],$$

define

$$s_{i,j} = \langle V_c(K), V_c(\Psi_{i,j}) \rangle,$$

and assemble

$$S = (s_{i,j}) \in \mathbb{R}^{(n_v n_z) \times n_h}.$$

The final CP matrix is

$$S = \frac{1}{6} \begin{bmatrix} 13 & 9.5 & 13 & 21.5 & 21 \\ 13 & 9.5 & 13 & 21.5 & 21 \\ 20 & 16.25 & 18 & 24.75 & 24.5 \\ 20 & 16.25 & 18 & 24.75 & 24.5 \\ 27.5 & 21.25 & 17.5 & 25 & 18 \\ 27.5 & 21.25 & 17.5 & 25 & 18 \\ 29.5 & 18 & 12.5 & 21.5 & 16.5 \\ 29.5 & 18 & 12.5 & 21.5 & 16.5 \end{bmatrix}. \quad (59)$$

8 Conclusion

This paper has presented a novel convolutional neural network framework based on the STP, which replaces the conventional inner product with a dimension-free STP inner product. By leveraging this cross-dimensional operation, we have developed a padding-free convolutional product (STP-CP) that performs convolution directly over valid data regions, eliminating the need for zero-padding or explicit mask propagation. The proposed approach has been shown to effectively handle irregular images, partially damaged data, and cases with mismatched receptive field and kernel sizes, thereby reducing artifacts and noise commonly introduced by padding in traditional CNNs.

Furthermore, a unified matrix formulation for STP-based convolution has been established, demonstrating the framework's scalability to higher-order signals such as three-dimensional volumetric data. This work provides a principled and mathematically rigorous theoretical foundation for the forward propagation of robust feature extraction in real-world scenarios involving incomplete or irregular data.

While this study has deliberately focused on the theoretical formulation and feasibility of the forward model, it opens several important avenues for future research. The immediate next steps include the following.

- **Implementation of the backward pass and training analysis.** A critical next step is the formal derivation and efficient implementation of the backpropagation algorithm for the STP-CP layer. This will enable a comprehensive investigation into the training dynamics, stability, and convergence behavior of STP-CNNs, especially concerning the spatial variability of receptive field dimensions.

- **Large-scale empirical validation.** Building on a trainable model, a thorough empirical evaluation is essential. This includes systematic comparisons with existing convolutional variants on public datasets to quantify gains in accuracy and robustness for tasks involving irregular data.

- **Computational efficiency and integration.** Future work will also focus on optimized implementations for practical deployment, a detailed analysis of computational efficiency on large-scale problems, and integration into mainstream deep-learning libraries for applications in domains like medical imaging and remote sensing.

The present work solidifies the forward propagation foundation, thereby setting the stage for these subsequent investigations into a fully trainable STP-CNN system.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 62350037).

References

- 1 Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proc IEEE*, 1998, 86: 2278–2324
- 2 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: *Proceedings of NeurIPS*, 2012
- 3 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of CVPR*, 2016. 770–778
- 4 Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Massachusetts: MIT Press, 2016
- 5 Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*, 2016, 529: 484–489
- 6 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. In: *Proceedings of ICLR*, 2016

- 7 Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks. In: Proceedings of ICCV, 2017. 764–773
- 8 Graham B. Submanifold sparse convolutional networks. ArXiv:1706.01307
- 9 Liu G, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions. In: Proceedings of ECCV, 2018. 85–100
- 10 Yu J, Lin Z, Yang J, et al. Free-form image inpainting with gated convolution. In: Proceedings of ICCV, 2019. 4470–4480
- 11 Chetlur S, Woolley C, Vandermersch P, et al. cuDNN: Efficient primitives for deep learning. ArXiv:1410.0759
- 12 Lavin A, Gray S. Fast algorithms for convolutional neural networks. In: Proceedings of CVPR, 2016. 4013–4021
- 13 Cheng D. From Dimension-Free Matrix Theory to Cross-Dimensional Dynamic Systems. London: Elsevier, 2019
- 14 Cheng D. On equivalence of matrices. *Asian J Math*, 2019, 23: 257–348
- 15 Cheng D. From DK-STP to non-square general linear algebra and general linear group. *Commun Inf Syst*, 2024, 24: 1–60
- 16 Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: Proceedings of MICCAI, 2016. 424–432
- 17 Kilmer M E, Braman K, Hao N, et al. Third-order tensors as operators on matrices: a theoretical and computational framework with applications in imaging. *SIAM J Matrix Anal Appl*, 2013, 34: 148–172
- 18 Lim L H. Tensors and hypermatrices. In: Handbook of Linear Algebra. 2nd ed. Boca Raton: CRC Press, 2013
- 19 Schaefer H H. Topological Vector Space. New York: Springer-Verlag, 1970
- 20 Edwards R E. Fourier Series. 2nd ed. New York: Springer-Verlag, 1979
- 21 Zhang A, Lipton Z C, Smola A J. Dive Into Deep Learning. Cambridge: Cambridge University Press, 2023
- 22 Wang Q, Yang Y, Jiang J. LEEDL Tutorial. Beijing: People's Posts and Telecommunications Press, 2024