PROBLEMS OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Sun’iy intellekt va mashinaviy o‘qitish muammolari

2026
Vol. 2, No. 1

Jurnal 2024-yilda ta’sis etilgan.
Yiliga 2 marta nashr qilinadi.

Ta’sischi:
Raqamli texnologiyalar va
sun’iy intellektni rivojlantirish
ilmiy-tadqiqot instituti

Bosh muharrir:
Mirzayev N.
Bosh muharrir o‘rinbosari:
Rajabov S.S., Axmedov D.D.
Mas’ul kotib:
Kaxarov Sh.S.
Tahririyat hay’ati:
Ablameyko S.V. (Belarus), Azamova N.A., Aydazade K.R. (Ozarbayjon), Andriyanov D.E. (RF), Atadjanov I.R., Bakayev I.I., Babayarov A.I., Bogush R.P. (Belarus), Daliyev Sh.K., Elov B.B., Eshonqulov X.I., Gupta A. (AQSH), Hindarto H. (Malayziya), Jiznyakov A.L. (RF), Juravkov M.A. (Belarus), Ignatyev N.A., Iordan V.I. (RF), Fazilov Sh.X., Filimonov N.B. (RF), Kalimoldayev M.N. (Qozog‘iston), Kamilov M.M., Karpenko A.P (RF), Kaxarov Sh.S., Madraximov Sh.F., Mamatov N.S., Mamyrbayev O.J. (Qozog‘iston), Mansurova M.E. (Qozog‘iston), Meliyev F.F., Merembayev T.J. (Qozog‘iston), Muxamedieva D.T., Muminov B.B., Musayev M.M., Nazirova E.Sh., Nuriddinov J.Z., Nuriyev U.G. (Ozarbayjon), Opanasenko V.N. (Ukraina), Palad V. (BB), Proletarskiy A.V. (RF), Rahim R. (Malayziya), Ravshanov N., Raxmatullayev M.A., Rustamov N. (Qozog‘iston), Rustamov S.S. (Ozarbayjon), Rzayev R.R. (Ozarbayjon), Sabziyev E.N. (Ozarbayjon), Starovoytov V.V. (Belarus), Subbotin S.A. (Ukraina), Tavboyev S.A., Tashev A. (Qozog‘iston), Tivari U.S., Tuzikov A.V. (Belarus), Urunboyev E., Xamdamov R.X., Yusupov O.R., Shafiyev T.R., Sharma G. (Hindiston).

Jurnal O‘zbekiston Respublikasi Prezidenti huzuridagi Axborot va ommaviy kommunikatsiyalar agentligi tomonidan ro‘yxatga olingan.
2024-yil 24-sentabrdagi
411962-son guvohnomasi.
ISSN 3093-9100.

Materiallarni qayta nashr etishda jurnalga havola qilish shart. Faktlarning aniqligi va ma’lumotlarning ishonchliligi uchun mualliflar mas’uldir.

Tahririyat manzili:
100125, Toshkent sh., Bo‘z-2 m-si, 17A
Тел.: +(998) 71 263-41-98
E-mail: paiml@airi.uz

Dizayn va sahifalash:
O‘rinboyev J.K.

RTSIR ITI bosmaxonasida chop etildi.
Bosma uchun ruxsat etilgan sanasi 24.06.2026.
Format 60x84 1/8.
Buyurtma No 1. Adadi 100 nusxa.

The journal was established in 2024.
2 issues are published per year.

Founder:
Digital Technologies and
Artificial Intelligence Development
Research Institute

Editor-in-chief:
Mirzaev N.
Deputy editor:
Radjabov S.S., Akhmedov D.D.
Executive secretary:
Kakharov Sh.S.
Editorial board members:

Ablameyko S.V. (Belarus), Azamova N.A., Aydazade K.R. (Azerbaijan), Andriyanov D.E. (Russia), Atadjanov I.R., Bakayev I.I., Babayarov A.I., Bogush R.P. (Belarus), Daliyev Sh.K., Elov B.B., Eshonkulov Kh.I., Gupta A. (USA), Hindarto H. (Malaysia), Jiznyakov A.L. (Russia), Zhuravkov M.A. (Belarus), Ignatyev N.A., Iordan V.I. (Russia), Fazilov Sh.Kh., Filimonov N.B. (Russia), Kalimoldayev M.N. (Kazakhstan), Kamilov M.M., Karpenko A.P. (Russia), Kakharov Sh.S., Madrakhimov Sh.F., Mamatov N.S., Mamyrbayev O.Zh. (Kazakhstan), Mansurova M.E. (Kazakhstan), Meliyev F.F., Merembayev T.Zh. (Kazakhstan), Mukhamedieva D.T., Muminov B.B., Musayev M.M., Nazirova E.Sh., Nuriddinov J.Z., Nuriyev U.G. (Azerbaijan), Opanasenko V.N. (Ukraine), Palad V. (United Kingdom), Proletarskiy A.V. (Russia), Rahim R. (Malaysia), Ravshanov N., Rakhmatullayev M.A., Rustamov N. (Kazakhstan), Rustamov S.S. (Azerbaijan), Rzayev R.R. (Azerbaijan), Sabziyev E.N. (Azerbaijan), Starovoytov V.V. (Belarus), Subbotin S.A. (Ukraine), Tavboyev S.A., Tashev A. (Kazakhstan), Tiwari U.S., Tuzikov A.V. (Belarus), Urunboyev E., Khamdamov R.Kh., Yusupov O.R., Shafiyev T.R., Sharma G. (India).

The journal is registered with the Agency of Information and Mass Communications under the President of the Republic of Uzbekistan.
Registration Certificate No. 411962
dated September 24, 2025.
ISSN 3093-9100.

When reproducing materials a reference to the journal is required. The authors are responsible for the accuracy of the facts and the reliability of the information.

Editorial office address:
17A, Buz-2, Tashkent, 100125 Uzbekistan
Tel.: +(99871)237-62-34
E-mail: paiml@airi.uz

Design and layout:
Urinboev J.K.

DTAIDRI printing office.
Signed for print 24.06.2026.
Format 60x84 1/8.
Order No. 1. Print run of 100 copies.

CONTENTS

MUNDARIJA

Fazilov Sh. Kh, Mirzaev N.M, Radjabov S.S., Mirzaev O.N, Tillavoldiev A.O. Pattern recognition: an analysis of the current stat	**4-21**	*Fazilov Sh. Kh, Mirzayev N.M, Radjabov S.S., Mirzayev O.N, Tillavoldiyev A.O.* Timsollarni tanib olish: zamonaviy holat tahlili
Mirzaev N.M., Rajabov S.S., Mirzaev O.N., Aripova S.O., Eshonkulova F.A. Model of recognition algorithms based on the identification of two-dimensional representative pseudo-objects in a training sample	**22-39**	*Mirzayev N.M., Rajabov S.S., Mirzayev O.N., Aripova S.O., Eshonkulova F.A.* O‘quv tanlanmasida ikki o'lchovli reprezentativ psevdoobyektlarni ajratishga asoslangan tanib olish algoritmlari modeli
Yusupov R.M, Sattarov Q.M, Abduraxmonov M.S Criteria for view formation and evaluation in multi-view analysis of multispectral images	**40-48**	*Yusupov R.M, Sattarov Q.M, Abduraxmonov M.S* Multispektral tasvirlarni ko‘p ko‘rinishli tahlil qilishda ko‘rinishlarni shakllantirish va baholash mezonlari
Yusupov O.R., Badalova L.B. An image enhancement algorithm for color videoendoscopic images	**49-61**	*Yusupov O.R., Badalova L.B* Rangli videoendoskopik tasvirlar sifatini yaxshilash algoritmi
Khandamov Y.Kh. Complementarity-based multi-level discriminative feature space construction	**62-75**	*Xandamov Y.X.* Komplementarlikka asoslangan ko‘p darajali diskriminativ belgilar fazosini shakllantirish
Yusupov O.R., Qodirov A.A. Evaluation of resnet models for palmprint identification using deep metric learning	**76-85**	*Yusupov O.R., Qodirov A.A.* Chuqur metrik o‘qitish asosida kaft izi identifikatsiyasi uchun resnet modellarini baholash
Mirzaeva G.R., Ibragimova S.N. Identification model based on the construction of threshold rules	**86-100**	*Mirzaeva G.R., Ibragimova S.N.* Bo‘sag‘a qoidalarini qurishga asoslangan identifikatsiya modeli
Salaev U.I. A joint neural model for morphological analysis and morpheme segmentation of uzbek words	**101-108**	*Salayev U.I.* O‘zbek so‘zlarining morfologik tahlili va morfemik segmentatsiyasi uchun qo‘shma neyron modeli
Yusupov O.R., Mikhliev R.R. Iris-based person identification: modern methods, challenges, and future directions	**109-118**	*Yusupov O.R., Mixliyev R.R.* Ko‘z rangdor pardasi asosida shaxsni identifikatsiyalash: zamonaviy usullar, muammolar va rivojlanish istiqbollari
Kakharov Sh.S, Najimov O., Meliboev A.I. LiteTabTransformer: *A* Pioneer of Lightweight Transformer Architecture for Efficient Network Intrusion Detection	**119-131**	*Kakharov Sh.S, Najimov O., Meliboev A.I.* LiteTabTransformer: tarmoqqa ruxsatsiz kirishlarni samarali aniqlash uchun yengil Transformer arxitekturasining ilg‘or model
Olimjonova S.G, Abdiyeva X. S. A classical image processing approach for left ventricular cavity segmentation in echocardiographic images	**132-142**	*Olimjonova S.G, Abdiyeva X. S.* Bell falaji va periferik yuz falajini sun’iy intellekt asosida avtomatlashtirilgan baholash: tizimli tahlil
Mamadjanov Sh.Sh. Artificial intelligence based automated assessment of bell’s palsy and peripheral facial paralysis: a systematic review	**143-152**	*Mamadjanov Sh.Sh.* Bell falaji va periferik yuz falajini sun’iy intellekt asosida avtomatlashtirilgan baholash: tizimli tahlil

© DTAID RI, 2026 **© RTSIR ITI, 2026**

UDC 004.93

COMPLEMENTARITY-BASED MULTI-LEVEL DISCRIMINATIVE FEATURE SPACE CONSTRUCTION

Khandamov Y.Kh.

*yxandamov@gmail.com
Digital Technologies and Artificial Intelligence Development Research Institute, Tashkent, Uzbekistan.

Abstract. Conventional multi-level feature fusion mechanisms in deep neural networks often fail to explicitly separate novel, complementary information from features already formed at preceding layers, introducing representational redundancy. This paper proposes a theoretical framework for Complementarity-Guided Multi-Level Discriminative Representation Learning (CLCD-RL). Representations extracted across network levels are mapped into a common space and decomposed into retained and complementary components via an orthogonal subspace projection mechanism. A task-aware gating function regulates residual contributions, while a supervised discriminative objective enforces intra-class compactness and inter-class separability. The theoretical and algebraic formulation establishes key conditions under which complementary features enhance discriminative power, providing a solid analytical foundation for hierarchical feature representations.

Keywords: deep learning, multi-level feature representation, discriminative learning, complementary features, orthogonal projection, feature redundancy, hierarchical representation, neural networks.

1 Introduction

The effectiveness of modern image recognition systems largely depends on their ability to construct informative and discriminative representations of input data. In conventional machine learning, feature extraction frequently relies on manually designed descriptors representing intensity distributions, textures, geometric structures, and statistical characteristics. Although such approaches provide interpretable information, their performance depends considerably on the quality of the selected descriptors and their correspondence to the classification problem.

Deep learning has substantially changed the feature extraction paradigm by enabling hierarchical representations to be learned directly from data. In convolutional neural networks, successive transformations construct increasingly abstract descriptions of input objects. Initial layers primarily capture local spatial structures, edges, and textures, while intermediate and deeper layers represent combinations of these structures and more complex semantic characteristics.

The development of residual neural networks demonstrated that deep hierarchical representations can be effectively learned through residual connections [1]. Similarly, Feature Pyramid Networks established the importance of combining representations extracted at different spatial resolutions [2]. These approaches demonstrated that information from different network levels may contribute to the recognition process.

However, the availability of multiple feature levels does not guarantee that their combination produces an optimal discriminative representation. Features extracted from different neural network layers frequently contain correlated or overlapping information. A feature represented at an intermediate level may partially reproduce information already available at a preceding level. Conversely, deeper layers may contain additional characteristics that improve the distinction between visually similar classes.

Conventional multi-level fusion mechanisms commonly employ concatenation, element-wise addition, weighted averaging, or attention-based integration. Although these operations provide

practical mechanisms for combining representations, they do not necessarily distinguish previously encoded information from complementary characteristics.

This limitation becomes particularly important when feature vectors have high dimensionality. Unrestricted combination of multiple representations may increase the number of parameters, produce unnecessary duplication, and complicate the optimization process. Furthermore, feature diversity alone does not imply discriminative usefulness. A component that differs from existing representations may describe irrelevant variations, such as illumination changes, background patterns, or noise.

Several approaches have been proposed to improve the quality of learned representations. Supervised contrastive learning encourages compactness among samples belonging to the same class while increasing separation between different classes [3]. Center loss introduces a compactness criterion into deep feature learning [4]. Barlow Twins [5] and VICReg [6] address statistical redundancy and representation collapse through covariance-related regularization mechanisms. Centered Kernel Alignment provides a means of investigating similarities between neural representations [7].

Despite these developments, an important methodological question remains: **how can new feature components be incorporated into a hierarchical representation while explicitly accounting for information already retained from preceding levels?**

The present research investigates this problem by introducing a complementarity-guided formulation of multi-level discriminative feature learning.

The principal objective is to develop a mathematical framework that combines hierarchical feature extraction, subspace-based redundancy analysis, complementary component formation, and supervised discriminative optimization.

The proposed contribution is not the invention of orthogonal projection, attention, or discriminative losses individually. Rather, it is their explicitly defined integration into a sequential feature construction procedure that distinguishes geometric complementarity from task-dependent discriminative usefulness.

2 Related Work and Problem Formulation

2.1 Multi-level feature representation in deep neural networks

Deep neural networks can be considered hierarchical transformation systems in which each layer forms a representation of the input data.

Let the labeled dataset be defined as

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N},$$

where $x_i \in \mathcal{X}$ denotes an input image, $y_i \in \{1, \dots, C\}$ represents its class label, N is the number of samples, and C is the number of classes.

Consider a deep neural network consisting of L feature extraction levels. The representation generated at the l-th level is

$$F_l(x) = E_l(F_{l-1}(x)), \qquad l = 1, \dots, L,$$

where E_l denotes the transformation performed by the corresponding neural network level and $F_0(x) = x$.

The feature tensors may have different spatial resolutions and channel dimensions. Consequently, direct mathematical comparison requires them to be transformed into a common representation space. Let P_l denote a pooling and projection operator associated with the l-th level. The aligned feature representation is defined as

$$h_l(x) = P_l\big(F_l(x)\big), \qquad h_l(x) \in \mathbb{R}^d \tag{1}$$

Here, d represents the dimensionality of the common feature space. The operator P_l may include spatial pooling, linear projection, normalization, or their combination. Its parameters are learned using the training data.

The resulting multi-level representation is described by the collection

$$\mathcal{H}(x) = \{h_1(x), h_2(x), \dots, h_L(x)\}.$$

A conventional concatenation strategy forms the combined representation as

$$z_{\text{concat}}(x) = [h_1(x)^\top, h_2(x)^\top, \dots, h_L(x)^\top]^\top.$$

Its dimensionality is Ld, assuming that every aligned vector contains d components. Alternatively, weighted fusion constructs

$$z_{\text{weighted}}(x) = \sum_{l=1}^{L} \alpha_l\,(x) h_l(x),$$

where $\alpha_l(x)$ represents a learned contribution coefficient.

Although these strategies combine information from several levels, they do not explicitly identify whether the information introduced at a particular level is already represented by previously retained feature directions. This motivates the introduction of a mechanism that distinguishes inherited and complementary components.

2.2. Discriminative representation learning

The objective of discriminative representation learning is to construct a feature space in which samples belonging to the same class are relatively compact, while samples from different classes remain sufficiently separated. The classical Fisher discriminant criterion formalizes this principle through within-class and between-class scatter [10]. In deep learning, related objectives have been implemented through center loss [4], supervised contrastive learning [3], and other embedding-based optimization methods.

Let $z_i \in \mathbb{R}^d$ denote the learned representation of sample x_i. The centroid of class c can be defined as

$$\mu_c = \frac{1}{N_c} \sum_{i:y_i=c} z_i,$$

where N_c denotes the number of samples belonging to class c.

The within-class scatter matrix is

$$S_W = \sum_{c=1}^{C} \sum_{i:y_i=c} (z_i - \mu_c)(z_i - \mu_c)^\top.$$

The between-class scatter matrix is

$$S_B = \sum_{c=1}^{C} N_c\,(\mu_c - \mu)(\mu_c - \mu)^\top,$$

where

$$\mu = \frac{1}{N} \sum_{i=1}^{N} z_i$$

denotes the overall feature centroid.

A discriminative representation should ideally exhibit relatively small within-class scatter and sufficiently large between-class scatter. However, optimizing discrimination alone does not explicitly address information duplication between different network levels.

2.3. Representation similarity and redundancy

The problem of redundancy arises when multiple feature components describe closely related characteristics of the same input object. Statistical redundancy may be investigated using correlation coefficients, covariance structures, or representation similarity measures. Barlow Twins introduces a cross-correlation-based objective that discourages redundant representation dimensions while preserving consistency between transformed views [5]. VICReg incorporates variance, invariance, and covariance regularization to support representation learning [6]. Centered Kernel Alignment,

introduced as a tool for comparing neural representations in [7], provides another perspective on representational similarity.

For centered representation matrices H_a and H_b, linear CKA may be written as

$$\mathrm{CKA}(H_a, H_b) = \frac{\|H_a^\top H_b\|_F^2}{\|H_a^\top H_a\|_F \|H_b^\top H_b\|_F}.$$

This quantity describes the similarity of the observed representation structures.

Nevertheless, similarity and task-level redundancy are not equivalent. Two feature representations may exhibit considerable correlation while still providing different information for classification. Likewise, statistically dissimilar representations are not necessarily discriminative.

Therefore, the proposed theoretical framework distinguishes three related concepts: representation similarity, geometric complementarity, and discriminative usefulness.

2.4. Mathematical formulation of the research problem

Let the aligned feature vectors extracted from different neural network levels be

$$h_1, h_2, \ldots, h_L \in \mathbb{R}^d.$$

The objective is to construct a final representation z_L that satisfies the following methodological requirements. First, the representation should preserve information useful for distinguishing classes. Second, when a new feature level is incorporated, its contribution should be evaluated relative to information already represented by preceding levels. Third, the feature construction procedure should remain numerically stable and compatible with gradient-based learning. The proposed solution is formulated as a sequential process in which each additional feature representation is decomposed into an inherited component and a complementary residual.

3 Complementarity-Guided Multi-Level Discriminative Feature Learning

3.1. Construction of the retained feature subspace

The proposed CLCD-RL framework begins by extracting aligned feature vectors from selected levels of a deep neural network. For the first level, the initial representation is defined as

$$z_1 = \phi(h_1),$$

where ϕ represents a normalization function.

For subsequent levels, information already represented by the preceding feature vectors is summarized through an orthonormal basis.

Let

$$B_{l-1} \in \mathbb{R}^{d \times r_{l-1}}$$

denote the basis associated with the previously retained feature subspace, where r_{l-1} is its rank.

The basis is required to satisfy

$$B_{l-1}^\top B_{l-1} = I_{r_{l-1}}.$$

For a particular input sample, the basis may be constructed by orthonormalizing previously accepted feature directions using QR decomposition or singular value decomposition. The corresponding orthogonal projection matrix is

$$\Pi_{l-1} = B_{l-1} B_{l-1}^\top. \quad (2)$$

The projection matrix satisfies

$$\Pi_{l-1}^\top = \Pi_{l-1}$$

and

$$\Pi_{l-1}^2 = \Pi_{l-1}.$$

These properties indicate that the operator is symmetric and idempotent. The projection represents the component of a feature vector belonging to the subspace already retained from preceding levels. Importantly, the retained basis must be constructed according to an explicitly defined rule. Orthogonalization by itself does not establish whether a feature direction contains useful class information.

3.2. Decomposition into inherited and complementary components

Consider a feature vector h_l extracted from the l-th network level. The component represented by the preceding feature subspace is

$$a_l = \Pi_{l-1} h_l.$$

The complementary component is defined as

$$r_l = (I_d - \Pi_{l-1}) h_l \tag{3}$$

where, I_d denotes the $d \times d$ identity matrix.

Substituting (2) into (3), we obtain

$$r_l = (I_d - B_{l-1} B_{l-1}^{\top}) h_l.$$

The original feature vector can consequently be decomposed as

$$h_l = a_l + r_l.$$

The inherited component a_l belongs to the retained subspace, whereas r_l is orthogonal to that subspace. This decomposition provides a mathematically explicit means of distinguishing feature directions that are represented by preceding levels from directions that are not contained in the selected retained subspace. The residual r_l is therefore interpreted as a geometrically complementary component.

However, its orthogonality does not guarantee statistical independence or class-related informativeness. A residual may also contain noise, background information, or other non-discriminative variations. For this reason, the proposed framework introduces a task-aware mechanism for controlling residual inclusion.

3.3. Task-aware selection of complementary features

Not every complementary residual should automatically contribute to the final representation. The usefulness of a residual depends on its relationship to the classification objective and to the already accumulated representation. To regulate this contribution, a learnable gating function is introduced.

Let $g_l \in (0,1)$ denote a scalar coefficient that controls the contribution of the residual generated at level l.

The gating function is defined as

$$g_l = \sigma\left(w_l^{\top} \begin{bmatrix} h_l \\ z_{l-1} \\ \| r_l \|_2 \end{bmatrix} + b_l \right). \tag{4}$$

Where w_l is the learnable parameter vector of the gating function; b_l is its bias; $\sigma(\cdot)$ denotes the logistic sigmoid function; $\| r_l \|_2$ is the Euclidean norm of the complementary residual; z_{l-1} is the accumulated representation obtained from preceding levels.

The coefficient g_l is learned through the overall supervised optimization objective. When g_l approaches zero, the corresponding residual contributes little to the updated representation.When g_l approaches one, the residual contribution is increased.

The gating function itself does not directly measure conditional mutual information or feature importance. Rather, it provides a trainable weighting mechanism whose usefulness must be established through the classification objective and subsequent analysis.

3.4. Sequential formation of the discriminative representation

The complementary residual is incorporated into the accumulated representation through an incremental update.

Let $W_l \in \mathbb{R}^{d \times d}$ denote a learnable feature transformation matrix. To preserve the orthogonality of the transformed residual relative to the previously retained subspace, define

$$\tilde{r}_l = (I_d - \Pi_{l-1}) W_l r_l.$$

The updated representation is

$$u_l = z_{l-1} + g_l \tilde{r}_l.$$

After normalization, the representation becomes

$$z_l = \frac{z_{l-1} + g_l \tilde{r}_l}{\|z_{l-1} + g_l \tilde{r}_l\|_2 + \varepsilon}. \quad (5)$$

Where, $\varepsilon > 0$ is a small constant introduced to ensure numerical stability.

The projected transformation is important because an unrestricted matrix W_l may rotate the residual back into the previously represented subspace. Applying the orthogonal complement projector after the transformation ensures that

$$B_{l-1}^{\top} \tilde{r}_l = 0.$$

When the retained basis contains z_{l-1}, the residual update is also orthogonal to the accumulated representation. The procedure is repeated for levels $l = 2, \ldots, L$.

Consequently, the final representation z_L is formed through the sequential incorporation of controlled complementary components. The proposed formulation differs from unrestricted concatenation because each new feature level is explicitly decomposed relative to previously retained directions before being incorporated. Nevertheless, the quality of the resulting representation remains dependent on the learned projections, retained rank, gate parameters, and supervised objective.

3.5. Discriminative optimization criterion

The proposed feature construction process must preserve class-related information. For a classification problem containing C classes, let

$$p_{ic} = q_\psi(z_{L,i})_c$$

denote the predicted probability that sample i belongs to class c.

The standard cross-entropy objective is defined as

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} \mathbb{1}(y_i = c) \log p_{ic}. \quad (6)$$

Where, $\mathbb{1}(\cdot)$ is the indicator function. Cross-entropy optimizes the relationship between the final learned representation and the predicted class probabilities. However, this loss does not explicitly require compact feature distributions within classes or large separation between class centroids. To address this issue, a discriminative regularization criterion is considered.

For the final representations $z_{L,i}$, define the within-class and between-class scatter matrices as described in Section 2.2. A normalized Fisher-type discriminative loss is then expressed as

$$\mathcal{L}_{\text{dis}} = \frac{\text{tr}(S_W)}{\text{tr}(S_B) + \varepsilon}.$$

The minimization of this ratio encourages a reduction in within-class scatter relative to between-class scatter. To prevent trivial scaling effects, this criterion is applied to normalized representations. The discriminative term is calculated only when sufficient class information is available. In particular, minibatches containing fewer than two represented classes do not provide a meaningful between-class comparison. The criterion may therefore be evaluated using class-balanced minibatches or training-only class prototypes maintained across optimization steps.

3.6. Complementarity regularization

Although orthogonal projection removes the component contained in the retained subspace, uncontrolled residual incorporation may increase representation complexity. A separate regularization mechanism is introduced to regulate the contribution of complementary feature directions.

Let

$$e_{i,l} = g_{i,l}^2 \parallel \tilde{r}_{i,l} \parallel_2^2$$

denote the squared magnitude of the accepted residual contribution for sample i at level l.

Define its mean value as

$$\overline{e} = \frac{1}{N(L-1)} \sum_{i=1}^{N} \sum_{l=2}^{L} e_{i,l}.$$

A bounded activity regularizer can be expressed as

$$\mathcal{L}_{\text{comp}} = [\max(0, e_{\min} - \overline{e})]^2 + [\max(0, \overline{e} - e_{\max})]^2. \quad (8)$$

Here, $e_{\min}$ and $e_{\max}$ are predefined bounds satisfying

$$0 \le e_{\min} < e_{\max}.$$

The purpose of this criterion is not to prove that residual energy corresponds to useful information. Instead, it controls the magnitude of accepted residual contributions and discourages two limiting situations: a representation in which all residual contributions disappear and one in which residual contributions become excessively large. The regularizer is zero when the mean accepted residual energy lies within the prescribed interval. The bounds must be selected using development data and should not be interpreted as universal constants.

If the projection mechanism is exact, an overlap penalty of the form $\| B_{l-1}^{\top} \tilde{r}_l \|_2^2$ would vanish identically. Such a term would therefore provide no additional optimization signal and is not included as an independent contribution to the proposed loss.

3.7. Stability of learned representations

In addition to discriminative quality, the learned representation should exhibit appropriate consistency under transformations that preserve the class of the input object. Let t_1 and t_2 denote two label-preserving transformations.

The corresponding representations are

$$z_i^{(1)} = z_L(t_1(x_i))$$

and

$$z_i^{(2)} = z_L(t_2(x_i)).$$

A representation consistency loss can be defined as

$$\mathcal{L}_{\text{stab}} = \frac{1}{N} \sum_{i=1}^{N} \left\| z_i^{(1)} - z_i^{(2)} \right\|_2^2.$$

The objective encourages representations obtained from two admissible transformations of the same object to remain close. However, the transformations must preserve the relevant class information. Otherwise, enforcing consistency may eliminate discriminative characteristics.

3.8. Combined optimization objective

The overall optimization criterion combines classification, discrimination, complementary contribution control, and representation consistency:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_{\text{dis}} \mathcal{L}_{\text{dis}} + \lambda_{\text{comp}} \mathcal{L}_{\text{comp}} + \lambda_{\text{stab}} \mathcal{L}_{\text{stab}}. \quad (9)$$

The coefficients

$$\lambda_{\text{dis}}, \lambda_{\text{comp}}, \lambda_{\text{stab}} \ge 0$$

regulate the relative influence of the corresponding objectives.

The parameters of the encoder, alignment operators, gating functions, residual transformations, and classifier are optimized according to

$$\Theta^* = \arg\min_{\Theta} \mathcal{L}_{\text{total}}(\Theta).$$

Here, Θ denotes the collection of trainable parameters.

The proposed optimization formulation provides a unified mechanism for forming hierarchical representations while distinguishing geometric complementarity from supervised discriminative objectives.

4 Theoretical Properties of the Proposed Approach

4.1. Orthogonality of complementary feature components

Proposition 1. Let $B \in \mathbb{R}^{d \times r}$ be a matrix with orthonormal columns satisfying

$$B^\top B = I_r.$$

For any feature vector $h \in \mathbb{R}^d$, define the complementary residual as

$$r = (I_d - BB^\top)h.$$

Then r is orthogonal to every column of B.

Proof.

Multiplying both sides by $B^\top$, we obtain

$$B^\top r = B^\top (I_d - BB^\top)h.$$

Expanding the expression gives

$$B^\top r = B^\top h - B^\top BB^\top h.$$

Since $B^\top B = I_r$,

$$B^\top r = B^\top h - I_r B^\top h = 0.$$

Therefore,

$$B^\top r = 0.$$

This proves that the residual lies in the orthogonal complement of the retained subspace.

The proposition guarantees geometric orthogonality under the chosen Euclidean metric. It does not guarantee statistical independence or discriminative relevance.

4.2. Energy decomposition of feature representations

Proposition 2. Under the assumptions of Proposition 1, the feature vector can be represented as the sum of two mutually orthogonal components:

$$h = BB^\top h + r.$$

Consequently,

$$\| h \|_2^2 = \| BB^\top h \|_2^2 + \| r \|_2^2.$$

Proof.

Let

$$a = BB^\top h.$$

From Proposition 1, $a^\top r = 0$.

Using $h = a + r$, we obtain

$$\begin{aligned} \| h \|_2^2 &= (a + r)^\top (a + r) \\ &= \| a \|_2^2 + 2a^\top r + \| r \|_2^2 \\ &= \| a \|_2^2 + \| r \|_2^2. \end{aligned}$$

Hence,

$$\| r \|_2^2 \leq \| h \|_2^2$$

and therefore

$$\| r \|_2 \leq \| h \|_2.$$

The obtained result demonstrates that orthogonal residual extraction does not increase the Euclidean norm of the original feature vector. This property applies to the projection operation itself. It does not imply that the entire neural network is non-expansive or Lipschitz continuous with constant one.

4.3. Orthogonal residuals and discriminative usefulness

Proposition 3. The existence of a nonzero orthogonal residual does not guarantee an improvement in classification performance.

Justification.

Suppose the class label Y is completely determined by an existing representation h_1. Assume that an additional feature vector h_2 contains independent noise unrelated to Y, and that h_2 is orthogonal to the retained subspace generated by h_1.

In this situation,

$$r_2 = h_2 \neq 0.$$

Nevertheless, the additional feature provides no new information about the target label.

Under the stated independence assumption,

$$I(Y; r_2 \mid h_1) = 0,$$

where $I(\cdot;\cdot|\cdot)$ denotes conditional mutual information. Thus, the residual is geometrically complementary but not conditionally informative for classification. This example demonstrates that subspace orthogonality alone cannot guarantee predictive improvement. The discriminative objective and task-aware gating mechanism are therefore necessary parts of the proposed formulation, although their inclusion does not itself guarantee better generalization.

4.4. Interpretation of the sequential representation update

Consider the unnormalized representation update

$$u_l = z_{l-1} + g_l \tilde{r}_l.$$

Assume that z_{l-1} belongs to the previously retained subspace and that

$$B_{l-1}^{\top} \tilde{r}_l = 0.$$

Then,

$$z_{l-1}^{\top} \tilde{r}_l = 0.$$

Consequently,

$$\| u_l \|_2^2 = \| z_{l-1} \|_2^2 + g_l^2 \| \tilde{r}_l \|_2^2.$$

This identity shows that the energy of the unnormalized updated representation can be decomposed into retained and accepted residual contributions when the orthogonality conditions hold. However, this equality should not be interpreted as proof that every residual improves feature quality. An increase in feature energy is not equivalent to an increase in class separability. The role of supervised optimization is to determine whether the additional representation directions are beneficial for the learning task.

4.5. Computational complexity

Let d denote the feature dimension and r_{l-1} the rank of the retained basis at level l.

The multiplication

$$B_{l-1}^{\top} h_l$$

requires approximately $O(dr_{l-1})$ arithmetic operations.

The subsequent reconstruction

$$B_{l-1}(B_{l-1}^{\top} h_l)$$

has the same asymptotic complexity.

Therefore, when the retained basis is already available, the cost of computing the complementary residual is

$$O(dr_{l-1}).$$

Across L levels, the projection-related cost is approximately

$$O\left(d \sum_{l=2}^{L} r_{l-1}\right).$$

This estimate excludes the cost of constructing and maintaining the orthonormal basis. Exact orthogonalization may introduce additional computational overhead, especially when feature dimensionality and retained rank are large. Low-rank approximations, truncated decompositions, or stabilized orthogonalization procedures may reduce this cost, but their effects on representation quality and gradient propagation require separate investigation.

5 Algorithmic Procedure and Methodological Verification

The proposed CLCD-RL method can be expressed as a sequential feature formation algorithm. The input consists of a labeled training dataset, a deep feature encoder, a set of selected representation levels, the common feature dimension, and optimization parameters. The output is a trained feature formation model and its corresponding discriminative representation.

Algorithm 1. Complementarity-Guided Multi-Level Discriminative Feature Representation Learning

1. Extract feature tensors from the selected encoder levels.
2. Transform each tensor into a common d-dimensional feature space using P_l.
3. Initialize the accumulated representation using the first feature level.
4. Construct an orthonormal basis from the previously retained feature directions.
5. For each subsequent level, calculate the inherited and orthogonal residual components.
6. Evaluate the task-aware gate and incorporate the projected residual into the accumulated representation.
7. Normalize the updated representation and continue until all selected levels have been processed.
8. Compute the classification, discriminative, complementarity, and stability losses.
9. Update trainable parameters using gradient-based optimization.
10. Return the trained model and final representation z_L.

The basis construction procedure requires particular attention. A sample-specific orthonormal basis and a basis estimated from training batches represent different mathematical formulations.
In the present theoretical model, the basis is defined for each sample using retained feature directions. This choice makes the orthogonality propositions explicit, but it introduces input-dependent geometry. A batch-level or globally learned subspace could provide a more stable population representation, although such an extension would require separate analysis.

The differentiation procedure must also be specified in any implementation. Exact differentiation through QR or SVD operations may become unstable when singular values coincide or when the retained rank changes. The use of detached basis updates may improve numerical behavior but changes the gradients of the optimization problem.

Therefore, differentiable orthogonalization, fixed-rank approximations, and detached subspace updates should be treated as distinct implementation variants.

5.1. Theoretical comparison with existing feature formation strategies

The principal characteristics of the proposed formulation can be summarized as follows.

Method	Multi-level features	Explicit subspace decomposition	Supervised discrimination	Conditional residual inclusion
Feature concatenation	Yes	No	Depends on training objective	No
Weighted feature fusion	Yes	No	Depends on training objective	Weighted, without explicit residual decomposition
Attention-based fusion	Yes	Generally no	Depends on training objective	Attention-based weighting
Orthogonal residual fusion	Yes	Yes	Not necessarily	Generally uncontrolled
Proposed CLCD-RL	Yes	Yes	Yes	Task-aware gating

This comparison describes methodological mechanisms rather than observed performance. In particular, some existing architectures may incorporate combinations of projection, attention, contrastive learning, and feature decorrelation. The comparison does not establish that such combinations are absent from the broader literature.

The distinguishing feature of the proposed formulation is its explicitly specified sequential interaction between retained-subspace construction, residual decomposition, supervised discrimination, and controlled residual incorporation.

5.2. Conditions for subsequent experimental verification

Although the present study focuses on theoretical development, the proposed mathematical framework generates several experimentally testable hypotheses.

First, the impact of residual decomposition should be evaluated against direct feature concatenation and weighted fusion under matched model capacities.

Second, the contribution of the discriminative objective should be isolated by comparing the complete formulation with a variant trained only using classification loss.

Third, the task-aware gating mechanism should be examined independently of the residual projection mechanism.

Fourth, the sensitivity of the method to retained subspace rank, feature dimension, and level ordering should be analyzed.

Finally, representation similarity and complementary information should be investigated using suitable diagnostic measures, including CKA and leave-one-level-out ablation.

Any future evaluation should employ independent training, validation, and test partitions. If the data contain repeated observations of the same object, patient, scene, or spatial region, the corresponding grouping structure must be preserved during data partitioning. This is important because inappropriate model selection and repeated use of evaluation data can produce optimistic performance estimates [11]. The experimental protocol should therefore distinguish between theoretical properties established by the mathematical formulation and practical benefits that must be demonstrated using independent data.

6 Discussion

The proposed CLCD-RL framework addresses the problem of constructing discriminative representations from multiple neural network levels by explicitly considering the relationship between previously retained features and newly extracted information. The main methodological principle is that hierarchical feature representations should not be combined without examining their overlapping components.

A deep feature representation may contain both inherited information and additional directions that are not represented by preceding levels. The proposed orthogonal decomposition makes this distinction mathematically explicit. However, the interpretation of the residual requires careful consideration.

Orthogonality is a geometric property defined relative to a selected feature subspace and coordinate system. It does not imply statistical independence, semantic novelty, or predictive usefulness. Similarly, the magnitude of a residual is not a universal measure of its contribution to classification. A large residual may contain irrelevant information, while a relatively small residual may describe an important class-specific characteristic.

Consequently, residual extraction is combined with a supervised discriminative objective and a trainable inclusion mechanism. The discriminative objective encourages compactness within classes and separation between different classes. The gating function controls the contribution of residual features during sequential representation construction.

These mechanisms serve different purposes. Orthogonal projection characterizes geometric complementarity, while supervised optimization determines how the constructed representation supports the classification task. Another important issue concerns the geometry of the common feature space. The projection operators P_l are trainable, and therefore the coordinates of the aligned representations depend on the learned transformations.

An orthogonal residual in one representation space need not retain the same interpretation after an arbitrary change of coordinates. The proposed decomposition is consequently meaningful only relative to the chosen embedding space, metric, and basis construction procedure.

The retained subspace rank also has a substantial influence on the method. If the retained basis spans the entire feature space, then

$$\Pi_{l-1} = I_d$$

and

$$r_l = 0.$$

In this case, no additional complementary direction can be incorporated from the current feature level. Conversely, if the retained rank is zero, then

$$\Pi_{l-1} = 0$$

and

$$r_l = h_l.$$

The mechanism then loses its ability to distinguish inherited and complementary directions.

These limiting cases demonstrate that rank selection is an essential component of the formulation. A fixed rank budget or training-based rank selection criterion may be used to control this property, although its optimal configuration cannot be determined theoretically for all datasets. The proposed framework also has limitations related to correlated but complementary information. Two strongly correlated features may provide useful conditional information for classification. Removing their overlapping geometric component does not necessarily preserve every relevant nonlinear relationship. Furthermore, nonlinear dependencies may persist between orthogonal residuals and retained representations. Thus, the proposed method should not be described as eliminating all statistical redundancy. It removes linear projections onto the chosen retained subspace under the adopted Euclidean geometry.

A possible theoretical extension is to replace Euclidean orthogonality with a generalized metric. Let $M \in \mathbb{R}^{d\times d}$ be a symmetric positive-definite matrix.

A weighted projection operator may be defined as

$$\Pi_M = B(B^\top MB)^{-1}B^\top M,$$

provided that the inverse exists.

Such an extension would allow the geometry of the feature space to account for a specified metric. Nevertheless, learning an appropriate metric introduces additional estimation and optimization challenges. Another direction concerns conditional information. Ideally, the usefulness of a new feature level could be characterized through

$$I(Y; h_l \mid H_{<l}),$$

where $H_{<l}$ denotes information retained from preceding levels.

This quantity reflects class-related information contributed by the new representation beyond what is already available. However, reliable estimation of conditional mutual information in high-dimensional neural representations is challenging.

The proposed method therefore employs geometric decomposition and supervised optimization as practical mathematical mechanisms rather than claiming to directly optimize conditional mutual information. The theoretical contribution of CLCD-RL lies in providing an integrated and explicitly defined framework for studying these relationships. Its practical superiority over conventional fusion techniques remains an empirical question and should be evaluated through controlled experiments.

7. Conclusion

This paper presented a theoretical framework for complementarity-guided multi-level discriminative feature representation learning. The proposed CLCD-RL approach considers hierarchical feature construction as a sequential process in which representations extracted from different neural network levels are mapped into a common space and decomposed into previously represented and complementary components.

An orthogonal subspace projection mechanism was formulated to identify feature directions not contained in the retained representation subspace. A task-aware gating mechanism was introduced to regulate the incorporation of these components, while a supervised discriminative criterion was included to encourage class-related representation quality.

The mathematical analysis established the orthogonality of residual components, the corresponding energy decomposition, and a bound on the norm of the extracted residual.

It was also demonstrated that geometric complementarity alone cannot guarantee improved classification performance or statistical independence.

The proposed formulation therefore distinguishes between the algebraic construction of complementary feature directions and their task-dependent discriminative usefulness.

The framework provides a theoretical basis for investigating hierarchical feature formation through subspace decomposition and supervised learning. Future research should examine its practical effectiveness using independent datasets, matched multi-level fusion baselines, component ablations, and analyses of rank sensitivity, computational complexity, and representation stability.

References

[1] He, K., Zhang, X., Ren, S., Sun, J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[2] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125. DOI: 10.1109/CVPR.2017.106.

[3] Khosla, P., Teterwak, P., Wang, C., et al. Supervised Contrastive Learning. *Advances in Neural Information Processing Systems*, Vol. 33, 2020, pp. 18661–18673.

[4] Wen, Y., Zhang, K., Li, Z., Qiao, Y. A Discriminative Feature Learning Approach for Deep Face Recognition. *Proceedings of the European Conference on Computer Vision*, 2016, pp. 499–515. DOI: 10.1007/978-3-319-46478-7_31.

[5] Zbontar, J., Jing, L., Misra, I., LeCun, Y., Deny, S. Barlow Twins: Self-Supervised Learning via Redundancy Reduction. *Proceedings of the 38th International Conference on Machine Learning*, PMLR, Vol. 139, 2021, pp. 12310–12320.

[6] Bardes, A., Ponce, J., LeCun, Y. VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. *International Conference on Learning Representations*, 2022.

[7] Kornblith, S., Norouzi, M., Lee, H., Hinton, G. Similarity of Neural Network Representations Revisited. *Proceedings of the 36th International Conference on Machine Learning*, PMLR, Vol. 97, 2019, pp. 3519–3529.

[8] Hu, J., Shen, L., Sun, G. Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141. DOI: 10.1109/CVPR.2018.00745.

[9] Woo, S., Park, J., Lee, J.-Y., Kweon, I. S. CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision*, 2018, pp. 3–19. DOI: 10.1007/978-3-030-01234-2_1.

[10] Fisher, R. A. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, Vol. 7, No. 2, 1936, pp. 179–188. DOI: 10.1111/j.1469-1809.1936.tb02137.x.

[11] Cawley, G. C., Talbot, N. L. C. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*, Vol. 11, 2010, pp. 2079–2107.

[12] Wang, T., Isola, P. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. *Proceedings of the 37th International Conference on Machine Learning*, PMLR, Vol. 119, 2020, pp. 9929–9939.

[13] Chen, T., Kornblith, S., Norouzi, M., Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. *Proceedings of the 37th International Conference on Machine Learning*, PMLR, Vol. 119, 2020, pp. 1597–1607.

[14] van den Oord, A., Li, Y., Vinyals, O. Representation Learning with Contrastive Predictive Coding. *arXiv preprint*, arXiv:1807.03748, 2018.

[15] Nogueira, S., Sechidis, K., Brown, G. On the Stability of Feature Selection Algorithms. *Journal of Machine Learning Research*, Vol. 18, No. 174, 2018, pp. 1–54.

KOMPLEMENTARLIKKA ASOSLANGAN KO‘P DARAJALI DISKRIMINATIV BELGILAR FAZOSINI SHAKLLANTIRISH

Xandamov Y.X.

*yxandamov@gmail.com
Raqamli texnologiyalar va sun’iy intellektni rivojlantirish ilmiy-tadqiqot instituti

Annotatsiya. Chuqur neyron tarmoqlarda an’anaviy ko‘p darajali belgilarni birlashtirish usullari tarmoqning oldingi qatlamlarida shakllangan axborotdan yangi, komplementar belgilarni to‘liq ajratib bera olmaydi, bu esa ortiqcha (redundant) komponentlar paydo bo‘lishiga olib keladi. Ushbu maqolada komplementarlikka asoslangan ko‘p darajali diskriminativ belgilar ifodasini o‘rganish (CLCD-RL) nazariy konsepsiyasi taklif etilgan. Yondashuv doirasida turli darajalardagi belgilar umumiy fazoga o‘tkaziladi va ortogonal qism-fazo proyeksiyalash mexanizmi yordamida avvalgi va komplementar komponentlarga ajratiladi. Residual komponentlar ulushi vazifaga mos geyt funksiyasi bilan boshqariladi, diskriminativ maqsad funksiya esa sinflararo maksimal ajralishni ta’minlaydi. Taqdim etilgan matematik apparat va nazariy tahlillar komplementar komponentlarning diskriminativ samaradorligini asoslab beradi hamda iyerarxik belgilarni shakllantirish uchun mustahkam poydevor yaratadi.

Kalit so‘zlar: chuqur o‘qitish, ko‘p darajali belgilar reprezentatsiyasi, diskriminativ o‘rganish, komplementar belgilar, ortogonal proyeksiya, belgilar ortiqchaligi, iyerarxik reprezentatsiya, neyron tarmoqlar.

ФОРМИРОВАНИЕ МНОГОУРОВНЕВОГО ДИСКРИМИНАТИВНОГО ПРОСТРАНСТВА ПРИЗНАКОВ НА ОСНОВЕ КОМПЛЕМЕНТАРНОСТИ

Хандамов Й.Х.

*yxandamov@gmail.com
Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта

Аннотация. Традиционные методы многоуровневого слияния признаков в глубоких нейронных сетях не разделяют явным образом новую комплементарную информацию от признаков предыдущих слоев, порождая избыточность представлений. В данной работе предложена теоретическая концепция обучения многоуровневым дискриминативным представлениям на основе комплементарности (CLCD-RL). В рамках подхода признаки разных уровней отображаются в единое пространство и декомпозируются на базовые и комплементарные компоненты с помощью механизма ортогонального проецирования в подпространство. Вклад остаточных компонентов регулируется адаптивной функцией вентилирования (gating), а оптимизационный дискриминативный критерий обеспечивает компактность внутри классов и разделимость между ними. Разработанный математический аппарат обосновывает условия эффективности комплементарных признаков и создает теоретическую основу для иерархических представлений.

Ключевые слова: глубокое обучение, многоуровневое представление признаков, дискриминативное обучение, комплементарные признаки, ортогональное проецирование, избыточность признаков, иерархическое представление, нейронные сети.