

INNOVATION SCIENCE AND TECHNOLOGY

ISSUE 9, PART I / SEPTEMBER 2026

 Acceptance of papers



**Acceptance of
papers**

Published monthly



Topics

economics,
technology, social
sciences



EDITOR-IN-CHIEF:

Mirzaliyev Sanjar Makhmatjon ugli

DEPUTY EDITOR-IN-CHIEF:

Makhmudov Nosir Makhmudovich
DSc., Prof., Academician

DEPUTY EDITOR-IN-CHIEF:

Ochilov Bobur Bakhtiyor ugli – Senior
lecturer at TSUI

THE SCIENTIFIC-POPULAR ELECTRONIC
JOURNAL **"INNOVATION SCIENCE AND
TECHNOLOGY"** HAS BEEN REGISTERED
UNDER THE NUMBER **C-5669633** BY THE
AGENCY FOR INFORMATION AND MASS
COMMUNICATIONS (AOKA) OF THE
REPUBLIC OF UZBEKISTAN, EFFECTIVE
FROM OCTOBER 9, 2024.

CONTACTS

Phone: **+998 50 737 87 88**

Website: <https://ist-journal.uz>

Email: innovationist2025@gmail.com

The scientific electronic journal "Innovation Science and Technology" has been included in the list of scientific publications recommended for the publication of main scientific results of dissertations for the award of PhD and DSc degrees in economics and technical sciences, in accordance with the Resolution No. 370 of the Presidium of the Higher Attestation Commission of the Republic of Uzbekistan, dated May 8, 2025.

Editorial board:



Sharipov Kongiratbay Avezimbetovich,
Doctor of Technical Sciences (DSc), Professor



Abdurakhmanova Gulnora Kalandarovna, Doctor of
Economic Sciences (DSc), Professor



Cham Tat Huei,
Doctor of Philosophy (PhD), Professor (Malaysia)



Muhammad Imran Sadiq
Doctor of Philosophy in Economics (PhD), Professor,
Malaysia



Ahmed Aziz Ismail
Doctor of Technical Sciences (DSc),
Professor (Egypt)



Lee Chin
Doctor of Philosophy in Economics (PhD), (Malaysia)



Asongu Simplice
Doctor of Philosophy in Economics (PhD), Cameroon



Rui Dang
Doctor of Chemistry (DSc), Professor, China



Zahoor Ahmed
Doctor of Philosophy in Economics (PhD), Turkey



Shujaat Abbas
Doctor of Philosophy in Economics (PhD), Russia



Tina A Coffelt
Doctor of Philosophy in Educational Sciences (PhD),
USA



Abdikarimova Dinara Rustamxanovna
Doctor of Economic Sciences (DSc), Professor

Kurbonbekova Mohichehra Turobjonovna
Doctor of Economic Sciences (DSc), Professor

Alimardonov Ilkhom Muzrabshokovich
Doctor of Economic Sciences (DSc), Professor



Razakova Barno Sayfieva
Doctor of Philosophy in Economics (PhD)



Khasanov Sarvar Ulugbek ugli
Doctor of Philosophy in Economics (PhD)



Kholikova Rukhsora Sanjarovna
Associate Professor (PhD)

CONTENTS

ECONOMIC ESSENCE AND CLASSIFICATION OF FINANCIAL ERRORS AND IRREGULARITIES IN BUDGET ORGANIZATIONS AND CRITERIA FOR THEIR INTERNAL AUDIT ASSESSMENT.....	8
Mulayev Farxod Alisherovich	
THE IMPACT OF FISCAL DISCIPLINE AND PUBLIC DEBT ON NATIONAL CURRENCY STABILITY: THEORETICAL APPROACHES AND MODELING	14
G'afurov Olimjon G'olib o'g'li	
THE ROLE OF REMOTE BANKING SERVICES IN ENHANCING FINANCIAL INCLUSION IN UZBEKISTAN	20
Sharipov Tohir Mirakhimovich	
THE IMPORTANCE OF STANDARDIZATION REQUIREMENTS IN THE ACCREDITATION OF NON-DESTRUCTIVE TESTING LABORATORIES.....	26
Ahmedov G'ayratjon Mahmudjonovich	
THEORETICAL ASPECTS OF ASSESSING THE IMPACT OF CORPORATE GOVERNANCE ON ECONOMIC EFFICIENCY IN CONSTRUCTION ENTERPRISES.....	32
Giyosov Ilkhom Karimovich	
FOREIGN EXPERIENCE IN STATE REGULATION OF SUSTAINABLE TOURISM DEVELOPMENT AND THE POSSIBILITIES OF ITS APPLICATION IN UZBEKISTAN.....	39
O'sarov Abdulla Davronqulovich	
THE CURRENT STATE AND DEVELOPMENT TRENDS OF THE SMALL BUSINESS SECTOR IN THE REPUBLIC AND ITS IMPACT ON HUMAN CAPITAL.....	46
Ulugmuradova Nodira Berdimuradovna	
INCREASING ECONOMIC EFFICIENCY BASED ON DIGITAL PLATFORMS IN THE IMPLEMENTATION OF ALTERNATIVE ENERGY TECHNOLOGIES (THE CASE OF RURAL AREAS).....	53
Ashurov Azizbek Ergash ugli	
WAYS TO IMPROVE THE QUALITY OF ASSETS AND STABILITY OF THE CREDIT PORTFOLIO IN COMMERCIAL BANKS	58
Baymuratova Zina Aqilbekovna	
MEASURING GREEN EFFICIENCY IN THE MINING INDUSTRY THROUGH AN INTEGRATED INDICATOR-BASED APPROACH FROM NAVOI MINING AND METALLURGICAL COMPANY.....	63
Kurbanova Mehriniso, Ergashev Hamidulla	
CORRELATION BETWEEN MACROECONOMIC INDICATORS AND TAX REVENUES.....	69
Khakimov Ulugbek Abdullaevich	
ENHANCING ECONOMIC EFFICIENCY AND FORECAST INDICATORS OF SCALING BIG DATA SOLUTIONS IN THE TOURISM INDUSTRY	74
E. I. Temirov	
IMPROVING THE EFFICIENCY OF STATE FINANCIAL SUPPORT FOR FARMS: A RESULTS- AND RISK-ORIENTED APPROACH.....	81
Rashidov Baxadir Yusupovich	
THE ROLE OF REMOTE BANKING SERVICES IN ENHANCING FINANCIAL INCLUSION IN UZBEKISTAN	86
Sharipov Tohir Mirakhimovich	
ORGANIZATIONAL AND LEGAL FOUNDATIONS FOR ESTABLISHING A PERSONNEL RESERVE FOR LOCAL GOVERNMENT AUTHORITIES IN UZBEKISTAN.....	92
G'aniyev Elyor Sobirjonovich	
METAEXTRACT: A HYBRID METADATA EXTRACTION SYSTEM COMBINING RULE-BASED AND STATISTICAL APPROACHES	100
Rashid Muhtorovich Turgunbaev	
DIRECTIONS FOR USING FINANCIAL ANALYSIS RESULTS IN THE DEVELOPMENT OF E-COMMERCE PLATFORMS IN UZBEKISTAN.....	105
Amankulova Mukhlisa Narsafar kizi	

IMPROVING THE ORGANISATIONAL-ECONOMIC MECHANISMS FOR ENHANCING THE PROFITABILITY OF ENTREPRENEURIAL ACTIVITY	111
Inobatov Abror Boshlarovich	
THE CURRENT STATE AND DYNAMICS OF FINANCING ENTREPRENEURIAL ACTIVITIES THROUGH BANK LOANS.....	117
Ergashev Otamurod Tashtemirovich	
IMPROVING THE METHODOLOGY FOR ASSESSING THE EFFICIENCY OF LAND TAX ADMINISTRATION FOR INDIVIDUALS	123
Ramazonov Shamil Ulugbek ogli	
IMPROVING THE EFFICIENCY OF STRATEGIC RESOURCE ALLOCATION IN BANKING ACTIVITIES THROUGH ARTIFICIAL INTELLIGENCE.....	130
Shukhrat Komilovich Roziboyev	
STRATEGIC DIRECTIONS FOR ENHANCING THE COMPETITIVE POTENTIAL OF ENTERPRISES THROUGH INNOVATION-DRIVEN DEVELOPMENT	138
Karima Samatovna Tashmukhamedova	
FUNCTIONAL FEATURES OF SEMANTIC PHENOMENA IN ENGLISH AND UZBEK METALLURGICAL TERMINOLOGY	143
O'tkirova Shahlo Asqar qizi	
DECENTRALIZED TECHNOLOGIES AND ARTIFICIAL INTELLIGENCE AS DRIVERS OF CROWDFUNDING EVOLUTION	151
Karimova Aziza A'zamiddin qizi	
THEORETICAL FOUNDATIONS FOR USING ARTIFICIAL INTELLIGENCE IN ASSESSING CORPORATE FINANCIAL STABILITY	159
Jokhongir Hasanovich Dagarov	
A MODEL FOR ASSESSING THE ECONOMIC EFFICIENCY OF INTRODUCING DIGITAL PLATFORMS	163
Nabiev Zafar Irkinovich	
CHARGING OF SYNTHETIC FIBERS AND IMPROVEMENT OF EMULSION.....	169
Arabov Jamoliddin Sadriddinovich	
ECONOMETRIC ANALYSIS OF FACTORS AFFECTING THE FINANCING OF INNOVATIVE ACTIVITIES OF TOURIST ENTERPRISES IN THE SAMARKAND REGION	177
Ruzibaeva Nargiza Khakimovna	
IMPLEMENTATION OF QUALIFIED PERSONNEL TRAINING CLUSTERS IN THE TOURISM SECTOR OF THE REGIONAL ECONOMY	181
Raxmatov Adxam Itolmasovich	
ASSESSING THE DIGITAL MATURITY OF ENTERPRISES AND IMPROVING THE MECHANISMS FOR STRENGTHENING THEIR COMPETITIVENESS: AN INDEX-BASED FRAMEWORK FOR UZBEK INDUSTRIAL AND SERVICE FIRMS	189
Kungratov Ilmurod Kuzibay ugli	
DIRECTIONS FOR DEVELOPING EMPLOYEE COMPETENCIES IN THE DIGITAL ECONOMY	197
Musaev Ozodjon Shavkat Ogli	
SPEAKER DIARIZATION OF KARAKALPAK SPEECH USING A PRETRAINED SORTFORMER MODEL UNDER LOW-RESOURCE CONDITIONS.....	202
Kudaybergenov Tangirbergen Kadirbergenovich	
ANALYSIS OF THERMAL AND ELECTRICAL PARAMETERS OF A PHOTOVOLTAIC-THERMAL ELECTROBIOMEMBRANE REACTOR IN WASTEWATER TREATMENT FOR MICROPOLLUTANT REMOVAL	208
Sevara U. Mirzayorova	
MODELING MAGNETOELASTIC VIBRATIONS OF COMPLEX-SHAPED TRANSVERSELY ISOTROPIC PLATES USING THE R-FUNCTION METHOD.....	212
Faxriddin Nuraliyev, Nuriddin Tojiyev, Bobir Adilov	

OPTIMAL CONDITIONS FOR EXTRACTION OF VALUABLE COMPONENTS FROM RICE HUSK.....	221
Umarov Faxriddin, Rahimov Baxodir, Blsmanov Ne'matullo, Ataxo'jayev Abdubositxo'ja	
IMPROVING THE ORGANIZATIONAL AND ECONOMIC MECHANISM FOR THE DEVELOPMENT OF THE SERVICES SECTOR: PROBLEMS AND PROSPECTIVE DIRECTIONS	227
Yo'ldoshev Shuxratmirza Tursunboy o'g'li	
INVESTIGATION OF THE KINETICS AND OPTIMAL CONDITIONS FOR THE PRODUCTION OF $FeCl_3$ COAGULANT FROM TEBINBULOQ ORES.....	235
Nurmurodov To'ikin, Oljayev Dilshod, Mansurova Shabnam	
ENVIRONMENTAL PROTECTION THROUGH RESOURCE-EFFICIENT WASTE UTILIZATION.....	241
Toirova Nuriya, Urinova Xulkar	
TECHNOLOGICAL AND DIGITAL DIVERSIFICATION IN INDUSTRIAL ENTERPRISES: DIRECTIONS FOR ADAPTING FOREIGN EXPERIENCES TO THE CONDITIONS OF UZBEKISTAN	245
Asadov Elyorbek Kandiayrovich	
A PROCESS-AWARE DATA PREPROCESSING AND TIME-DELAY IDENTIFICATION FRAMEWORK FOR CEMENT CLINKER ENERGY MONITORING	250
Abzor Turakulov, Maksadkhon Yakubov, Zafar Turakulov	
DEVELOPMENT OF KPIS TO MEASURE THE SUCCESS OF INNOVATIVE ENVIRONMENTAL PROJECTS WITHIN THE COMPANY.....	260
Musadjanova Nargiza Abduvoxid qizi	
ANALYSIS OF THE EFFICIENT USE OF FINANCIAL RESOURCES IN ENSURING EFFECTIVE ENTREPRENEURIAL ACTIVITY IN THE REGION (ANALYSIS OF THE STATE OF ENTREPRENEURSHIP DEVELOPMENT IN FERGANA REGION).....	266
Bozorov Ilyos Isomiddinovich	
TIJORAT BANKLARIDA ICHKI AUDIT AMALLARINI BAJARISH.....	271
Yuldashev Xurshidbek Mamirjonovich	
THE ROLE OF FASHION ELEMENTS IN THE FORMATION OF A NATION BRAND: A CONCEPTUAL FRAMEWORK AND RESEARCH METHODOLOGY	277
Yusupova Mokhinur Farkhod qizi	
THE ORGANISATIONAL AND ECONOMIC MECHANISM FOR EVALUATING THE EFFECTIVENESS OF MARKETING PROMOTION OF EDUCATIONAL SERVICES.....	287
Umarov Shuhrat Bahodir o'g'li	
COMPUTER VISION TECHNOLOGIES TO IMPROVE THE IMAGE QUALITY OF THE IRIS	297
Makhmudova Shakhzoda, Sharopova Barno, Saydazimov Javlonbek	
SPECIFIC ASPECTS OF ENHANCING INDUSTRIAL PRODUCTION COMPETITIVENESS.....	301
Olimov Maksudjon, Habibullayev Oybek	
COMPARATIVE EVALUATION OF CNN-LSTM AND SQUEEZEFORMER-TRANSFORMER ARCHITECTURES FOR REAL-TIME UZBEK SIGN LANGUAGE RECOGNITION	307
Jasurbek Ergashev, Munira Rakhmatova, Mukhtor Eshimov, Ibragim Atadjonov, Sajid Khan	
ECONOMIC CONTRIBUTION OF DOMESTIC TOURISM IN UZBEKISTAN.....	316
Abdurakhmonova Shakhzoda	
A METHODOLOGY FOR A MULTIDIMENSIONAL INTEGRAL IN-DEX OF DEMAND-PRODUCTION ALIGNMENT IN THE BUILDING MA-TERIALS INDUSTRY.....	321
Turdiev Abdullo Sagdullaevich	
THE IMPACT OF STATE BUDGET EXPENDITURES ON HUMAN CAPITAL DEVELOPMENT: INTERNATIONAL TRENDS, FISCAL EFFICIENCY, AND PRIORITY AREAS FOR UZBEKISTAN	328
Durdona Davletova	
FINANCIAL INFORMATION IN PUBLIC HEALTHCARE ORGANISATIONS: DEMAND, REGULATION AND THE WORKING CAPACITY OF THE ACCOUNTING METHOD	342
Kuliboyev Azamat Shonazarovich	
COMPARING ENCODER FUSION STRATEGIES IN HYBRID CNN-TRANSFORMER NETWORKS FOR 3D BRAIN TUMOR SEGMENTATION: A CONTROLLED ABLATION ON BRATS 2021.....	354
Abdullayev Temurbek, Gulbahor Abdukadirova	

COMPARING ENCODER FUSION STRATEGIES IN HYBRID CNN-TRANSFORMER NETWORKS FOR 3D BRAIN TUMOR SEGMENTATION: A CONTROLLED ABLATION ON BRATS 2021

Abdullayev Temurbek

PhD, Associate Professor, Department of Computer Engineering and Artificial Intelligence, Fergana State Technical University

temurbekm84@gmail.com

ORCID: 0000-0001-5946-7607

Gulbahor Abdukadirova

Department of Computer Engineering and Artificial Intelligence, Fergana State Technical University

nirvanath369@gmail.com

ORCID: 0009-0005-2035-2008

ABSTRACT. Hybrid CNN–Transformer architectures have become the dominant paradigm for volumetric medical image segmentation, combining the local inductive bias of convolutions with the long-range context modeling of self-attention. However, the point at which the two encoder streams are fused and the mechanism used to fuse them are rarely studied in isolation: most published gains conflate fusion design with unrelated changes in depth, patch size, or training schedule. This paper presents a controlled ablation that isolates the fusion strategy as the sole independent variable. Using a shared CNN and Swin-Transformer backbone, we compare six fusion strategies—early, late, additive, concatenation, cross-attention, and a proposed gated multi-scale cross-attention fusion—on the BraTS 2021 dataset (1,251 training cases) under identical training conditions. The proposed gated multi-scale cross-attention fusion achieves the best overall performance, with mean Dice scores of 94.2%, 89.1%, and 82.5% for the whole tumor (WT), tumor core (TC), and enhancing tumor (ET) sub-regions, respectively, and the lowest HD95 of 4.68 mm, outperforming the strongest single-branch baseline by 2.4–4.1 Dice points while adding only 12 million parameters.

KEYWORDS. Brain tumor segmentation; BraTS 2021; CNN–Transformer fusion; hybrid encoder; cross-attention; medical image segmentation; deep learning.

ANNOTATSIYA. Gibrir CNN–Transformer arxitekturalari hajmli tibbiy tasvirlarni segmentatsiyalashda yetakchi yondashuvga aylandi, chunki ular konvolyutsiyalarning lokal induktiv xususiyatlarini self-attention mexanizmining uzoq masofali kontekstni modellash tirish imkoniyatlari bilan birlashtiradi. Biroq, ikki enkoder oqimini qaysi bosqichda birlashtirish va ularni qanday mexanizm orqali integratsiya qilish masalalari kamdan-kam hollarda alohida o'rganiladi: chop etilgan ko'plab natijalarda fusion dizaynining ta'siri chuqurlik, patch o'lchami yoki o'qitish jadvalidagi boshqa o'zgarishlar bilan birgalikda baholanadi. Ushbu maqolada fusion strategiyasini yagona mustaqil o'zgaruvchi sifatida ajratib ko'rsatadigan nazorat qilinadigan ablyatsion tahlil taqdim etiladi. Umumiy CNN va Swin-Transformer tayanch arxitekturasidan foydalanish holda, biz BraTS 2021 ma'lumotlar to'plamida (1 251 ta o'qitish holati) bir xil o'qitish sharoitlarida oltita fusion strategiyasini — erta, kechki, additiv, konkatenatsiya, cross-attention hamda taklif etilgan gated multi-scale cross-attention fusion usullarini — taqqoslaymiz. Taklif etilgan gated multi-scale cross-attention fusion eng yaxshi umumiy natijani ko'rsatib, mos ravishda butun o'sma (WT), o'sma yadrosi (TC) va kontrastni kuchaytiruvchi o'sma (ET) subregionlari uchun o'rtacha Dice ko'rsatkichlarini 94,2%, 89,1% va 82,5% darajasida ta'minladi, shuningdek, eng past HD95 qiymati 4,68 mm ni tashkil etdi. Ushbu yondashuv eng kuchli bir tarmoqli bazaviy modeldan Dice ko'rsatkichi bo'yicha 2,4–4,1 punktga yuqori natija ko'rsatdi va bunda modelga atigi 12 million parametr qo'shildi.

KALIT SO'ZLAR. Bosh miya o'smasini segmentatsiyalash; BraTS 2021; CNN–Transformer fusion; gibrir enkoder; cross-attention; tibbiy tasvirlarni segmentatsiyalash; chuqur o'rganish.

АННОТАЦИЯ. Гибридные архитектуры CNN–Transformer стали ведущей парадигмой в области сегментации объёмных медицинских изображений, объединяя локальное индуктивное смещение свёрточных сетей с возможностями механизма self-attention по моделированию дальних контекстных зависимостей. Однако вопрос о том, на каком этапе следует объединять два потока энкодера и какой механизм использовать для их слияния, редко рассматривается изолированно: во многих опубликованных работах влияние стратегии слияния смешивается с другими изменениями, связанными с глубиной модели, размером патча или расписанием обучения. В данной

статье представлено контролируемое абляционное исследование, в котором стратегия слияния выделена в качестве единственной независимой переменной. Используя общую архитектурную основу CNN и Swin-Transformer, мы сравниваем шесть стратегий слияния — раннее, позднее, аддитивное, конкатенацию, cross-attention и предложенное gated multi-scale cross-attention fusion — на наборе данных BraTS 2021 (1 251 обучающий случай) при идентичных условиях обучения. Предложенный метод gated multi-scale cross-attention fusion продемонстрировал наилучшую общую эффективность: средние значения Dice составили 94,2%, 89,1% и 82,5% соответственно для всей опухоли (WT), ядра опухоли (TC) и контрастно-усиливающейся опухоли (ET), а минимальное значение HD95 составило 4,68 мм. Данный подход превзошёл наиболее сильную одноветвевую базовую модель на 2,4–4,1 пункта Dice, добавив при этом всего 12 миллионов параметров.

КЛЮЧЕВЫЕ СЛОВА. Сегментация опухолей головного мозга; BraTS 2021; CNN–Transformer fusion; гибридный энкодер; cross-attention; сегментация медицинских изображений; глубокое обучение.

INTRODUCTION

Accurate delineation of gliomas from multi-modal magnetic resonance imaging (MRI) is a prerequisite for treatment planning, surgical guidance, and longitudinal monitoring of brain tumor patients. Manual segmentation of the whole tumor, tumor core, and enhancing tumor sub-regions is time-consuming and suffers from inter-rater variability, which is why deep convolutional networks, and in particular the U-Net family, have become the standard tool for this task, with the BraTS challenge serving as the primary benchmark [1], [2].

Convolutional networks are well suited to this problem because of their strong local inductive bias, but they struggle to model long-range spatial dependencies for example, the relationship between a tumor's necrotic core and its distant peritumoral edema — without stacking many layers. Vision Transformers (ViT) and their hierarchical variants, most notably the Swin Transformer, address this limitation through self-attention, which gives each layer an effectively large receptive field from the first layer onward. This property has motivated a wave of hybrid CNN-Transformer segmentation networks, including TransUNet, UNETR, Swin UNETR, and nnFormer, that pair a convolutional stem or decoder with a Transformer-based encoder [3],[6].

A close reading of this literature reveals an important gap: the specific mechanism used to combine CNN and Transformer features— concatenation, addition, gating, or cross-attention, and whether it occurs once at the bottleneck or repeatedly across multiple encoder scales — is rarely the controlled variable in published comparisons. Papers introducing a new fusion mechanism typically also change backbone depth, patch embedding strategy, or training budget relative to the baselines they compare against, making it difficult to attribute reported gains specifically to the fusion design [7].

This paper addresses that gap through a controlled ablation study. We hold the CNN branch, the Transformer branch, the training data, the loss function, the optimizer, and the training budget fixed, and vary only the fusion strategy connecting the two encoder streams. Our contributions are threefold: (1) we implement and evaluate six encoder fusion strategies within a single shared codebase; (2) we propose a gated multi-scale cross-attention fusion module that adaptively weights CNN and Transformer contributions at each encoder resolution; (3) we provide a quantitative analysis of the accuracy-complexity trade-off across fusion strategies on BraTS 2021.

LITERATURE REVIEW

Research on encoder design for volumetric medical image segmentation can be grouped into three strands: convolutional baselines, Transformer-based baselines, and the hybrid CNN-Transformer architectures that motivate the present ablation.

Ronneberger et al. introduced the U-Net architecture, establishing the encoder–decoder-with-skip-connections template that remains the backbone of nearly all subsequent segmentation networks [9]. Building on this template, Isensee et al. showed with nnU-Net that a carefully engineered, self-configuring convolutional pipeline can match or exceed more architecturally novel methods on the BraTS benchmark [1]; the authors of the present paper treat this result as an important caution against attributing performance gains to architecture alone rather than to training protocol, and it is precisely this confound that the controlled design of the present study is intended to remove. Menze et al. established the BraTS challenge as the standard benchmark for comparing segmentation methods, and Baid et al. later extended it to the 2021 release used here [2], [8].

Vaswani et al. introduced the self-attention mechanism underlying all Transformer-based vision models [10], and Dosovitskiy et al. demonstrated that a pure Transformer, given sufficient data, can outperform convolutional networks on image classification [10]. Liu et al.'s Swin Transformer restricted self-attention to shifted local windows, making the mechanism computationally tractable for dense prediction and providing the backbone this paper adopts for its Transformer branch [7]. Chen et al. were among the first to combine the two

paradigms for medical segmentation, inserting a Transformer bottleneck between a CNN encoder and decoder in TransUNet [3]; the present authors note, however, that TransUNet's Transformer operates only at the lowest resolution, which limits information exchange with the CNN branch at other scales a limitation the fusion module proposed in this paper is designed to address. Hatamizadeh et al. extended the hybrid idea to full 3D volumes with UNETR [4], and subsequently replaced the plain ViT encoder with a hierarchical Swin encoder in Swin UNETR, reducing computational cost while preserving multi-scale features [5]; this paper's shared backbone follows the same Swin-based design, but, unlike Swin UNETR, treats the fusion mechanism, rather than the encoder itself, as the sole variable under study. Zhou et al.'s nnFormer instead interleaved convolution and windowed self-attention within a single encoder path rather than keeping two separate branches [6]; the present authors regard this single-stream interleaving and the dual-branch fusion studied here as complementary design philosophies rather than competing ones, since both address the same underlying need to combine local and global representations. Hu et al.'s squeeze-and-excitation gating mechanism, originally proposed for image classification, is adapted in the present paper's proposed fusion module to weight CNN and Transformer contributions per channel [10].

A recurring observation across this body of literature, and the central motivation for the present study, is that new fusion mechanisms are almost never isolated as the sole independent variable: published comparisons typically vary backbone depth, patch embedding, or training budget alongside the fusion design, which makes it difficult to attribute reported gains specifically to fusion [7].

Uzbek researchers have likewise addressed neural-network-based image segmentation, though largely outside the medical CNN-Transformer fusion setting examined here. Kurbanov and Beknazarova reviewed deep learning approaches to biomedical image segmentation, tracing the shift from classical convolutional encoder-decoder models toward more automated pipelines and concluding that architecture choice is only one of several determinants of segmentation quality [9]; the authors of the present paper share this conclusion and extend it specifically to the fusion sub-problem within hybrid encoders, which Kurbanov and Beknazarova's review does not address. Mamatov et al. compared U-Net, Fully Convolutional Networks, and Mask R-CNN for face-image segmentation, systematically weighing each architecture's accuracy, computational cost, and domain of application [10]; while their comparison, like much of the domestic literature, remains at the level of whole-architecture selection rather than encoder-fusion design, its methodology of controlled, criterion-based comparison across models directly informs the ablation protocol adopted in this paper, and its reported accuracy-cost trade-offs are consistent with the trends observed in the present results. Taken together, the domestic literature confirms that convolutional segmentation networks are an active and maturing research direction in Uzbekistan, while also showing that the specific question addressed here how CNN and Transformer streams should be fused within a single hybrid encoder has not yet been taken up locally, which is one of the gaps this paper aims to help close.

Read together, the foreign literature establishes that hybrid CNN-Transformer encoders outperform single-branch baselines, while the domestic literature confirms that convolutional segmentation networks are actively studied locally but that hybrid fusion design remains largely unexplored. The present study positions itself at this intersection: it adopts a foreign-benchmark backbone (Swin Transformer, BraTS 2021) while treating fusion strategy, isolated from all confounding factors, as its central object of study a controlled comparison design that, to the authors' knowledge, has not previously been reported in either body of literature.

RESEARCH METHODOLOGY

Related work is reviewed in detail in the Literature Review section above; the remainder of this section describes the experimental setup that follows from it.

Dataset. Experiments use the BraTS 2021 dataset, comprising 1,251 labeled training cases with four co-registered, skull-stripped MRI modalities (T1, T1ce, T2, FLAIR), each 240×240×155 voxels [8], [9]. Ground-truth labels distinguish the enhancing tumor (ET), tumor core (TC), and whole tumor (WT) sub-regions used for evaluation.

Shared backbone. All models share a common dual-branch encoder skeleton. The CNN branch is a 3D residual encoder with four downsampling stages (32–64–128–256 channels). The Transformer branch is a hierarchical Swin-style encoder [5] operating at the same four resolutions with matched channel widths. Both branches receive the same four-channel input, and the decoder is a symmetric convolutional upsampling path, identical across all conditions. Only the fusion module connecting the two encoder branches, and the corresponding decoder skip pathways, differ between conditions.

Fusion strategies compared. Eight configurations were evaluated: (i) CNN-only and (ii) Transformer-only single-branch baselines; (iii) early (input-level) fusion, where CNN and Transformer patch embeddings are concatenated at a shared stem; (iv) late (decision-level) fusion, where independent branch logits are

averaged just before decoding; (v) additive skip fusion, summing CNN and Transformer feature maps at each of four resolutions; (vi) concatenation fusion, concatenating features channel-wise at each resolution followed by a $1 \times 1 \times 1$ convolution; (vii) cross-attention fusion, applying bidirectional multi-head cross-attention at the bottleneck and two coarsest skip resolutions; and (viii) the proposed gated multi-scale cross-attention fusion, which extends cross-attention to all four resolutions and adds a lightweight per-channel gating branch [10] that adaptively weights the CNN and Transformer contribution at each scale before the fused features reach the decoder.

Training configuration. All eight configurations were trained from random initialization under an identical protocol: 128^3 randomly cropped patches with foreground oversampling, random flip/rotation/elastic-deformation augmentation, a combined Dice and focal cross-entropy loss (0.5/0.5), the AdamW optimizer (initial learning rate 1×10^{-4} , cosine annealing), 300 epochs with early stopping, and 5-fold cross-validation on the 1,251 labeled training cases [10]. Performance was evaluated using the Dice similarity coefficient and the 95th-percentile Hausdorff distance (HD95) for each of the three sub-regions.

ANALYSIS AND RESULTS

Table 1 reports Dice and HD95 scores for all eight configurations, averaged over the 5-fold cross-validation protocol. Both single-branch baselines underperform every fusion strategy: the CNN-only baseline achieves a mean WT Dice of 91.8%, while the Transformer-only baseline trails at 89.6%. Fusion strategies that interact only once (early and late fusion) yield the smallest improvements over the CNN-only baseline (+0.6 and +0.3 WT Dice), while strategies that interact repeatedly across multiple encoder scales show progressively larger gains. The proposed gated multi-scale cross-attention fusion achieves the best results across all three sub-regions and both metrics: mean Dice of 94.2% (WT), 89.1% (TC), and 82.5% (ET), and the lowest HD95 of 4.68 mm.

Table 1. Segmentation performance and model size across fusion strategies (mean over 5-fold cross-validation).

Fusion Strategy	Fusion Point	Dice WT	Dice TC	Dice ET	HD95, mm	Params, M
CNN-only (baseline)	—	91.8	85.2	78.4	6.91	18.2
Transformer-only	—	89.6	82.7	75.1	8.34	31.5
Early fusion	Stem	92.4	86.0	79.0	6.42	24.6
Late fusion	Output	92.1	85.8	78.9	6.58	26.1
Additive skip fusion	Enc-dec skips	92.9	86.9	80.1	5.87	27.4
Concatenation fusion	Bottleneck	93.0	87.1	80.3	5.79	28.0
Cross-attention fusion	Bottleneck+skips	93.7	88.4	81.6	5.12	29.8
Gated cross-attention (proposed)	Multi-scale	94.2	89.1	82.5	4.68	30.3

Source: developed by the author.

The gap between the best fusion strategy and the CNN-only baseline is largest for the enhancing tumor sub-region (+4.1 Dice points) and smallest for the whole tumor (+2.4 Dice points), consistent with ET being the hardest sub-region to segment. Comparing concatenation fusion and cross-attention fusion, which share the same fusion locations, cross-attention adds 0.7–1.3 Dice points at the cost of roughly 1.8 million additional parameters, indicating that learned, content-dependent weighting captures information that fixed-weight channel mixing does not. In terms of computational cost, moving from the CNN-only baseline (18.2M parameters) to the proposed fusion strategy (30.3M parameters) roughly doubles the parameter count, while inference time on a single A100 GPU increased from 1.8 to 3.1 seconds per full-volume prediction.

The results support the central hypothesis of this study: the choice of fusion strategy has a measurable and reproducible effect on segmentation accuracy that is separable from backbone capacity, since all eight configurations share identical CNN and Transformer branches, training data, and optimization settings. Two general design principles emerge. First, fusing complementary representations at multiple spatial resolutions, rather than only at the bottleneck, allows each decoder stage to draw on locally appropriate combinations of fine-grained and contextual information. Second, content-adaptive fusion mechanisms (cross-attention, gating) outperform fixed-weight mechanisms (addition, concatenation) by a consistent margin, indicating that the relative usefulness of CNN versus Transformer features varies across spatial location and tumor sub-region.

These findings should be interpreted in light of several limitations. All models were trained from random initialization without large-scale pretraining, which may understate the performance of the Transformer branch and fusion strategies that depend more heavily on it. Second, the ablation is restricted to encoder-side fusion;

decoder-side fusion remains a natural extension. Third, evaluation is limited to internal 5-fold cross-validation on BraTS 2021 rather than an independent external test set, so absolute figures should be interpreted as internally consistent rather than directly comparable to other groups' leaderboard results.

CONCLUSION AND SUGGESTIONS

This paper presented a controlled ablation study isolating the effect of encoder fusion strategy on 3D brain tumor segmentation accuracy within hybrid CNN–Transformer networks. By holding backbone architecture, training data, and optimization protocol fixed across eight configurations evaluated on BraTS 2021, we showed that fusion design alone accounts for a consistent range of performance, from single-branch baselines through single-point fusion to multi-scale, content-adaptive fusion. The proposed gated multi-scale cross-attention fusion achieved the best results across all three BraTS sub-regions and both evaluation metrics, improving mean Dice by 2.4–4.1 points over the strongest single-branch baseline while adding only 12 million parameters — a modest computational cost for a consistent and reproducible accuracy gain.

These findings carry a practical implication for practitioners designing hybrid segmentation networks: where the two encoder streams are fused matters as much as, and at times more than, which specific attention or convolutional blocks each branch contains. Fusing complementary representations at multiple spatial resolutions rather than only at the bottleneck, and doing so through content-adaptive mechanisms rather than fixed-weight addition or concatenation, offers a reliable, architecture-agnostic route to improved accuracy that does not require redesigning either encoder branch from scratch.

Several directions remain open. Because all eight configurations were trained from random initialization, future work should evaluate whether the relative ranking of fusion strategies persists when the Transformer branch is initialized from large-scale pretraining, which may narrow or widen the gap between mechanisms that depend more heavily on the Transformer stream. The present ablation is also restricted to encoder-side fusion; extending the same controlled methodology to decoder-side and multi-stage fusion is a natural next step. Finally, validating the proposed gated multi-scale cross-attention fusion on independent glioma cohorts and on segmentation benchmarks beyond BraTS 2021 would help establish whether the accuracy gains observed here generalize across imaging protocols and institutions, and whether they translate into clinically meaningful improvements in tumor delineation.

REFERENCES

1. · Isensee F., Jaeger P.F., Kohl S.A.A., Petersen J., Maier-Hein K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation // *Nature Methods*. 2021. Vol. 18, No. 2. P. 203–211. DOI: 10.1038/s41592-020-01008-z. <https://doi.org/10.1038/s41592-020-01008-z>
2. · Menze B.H., Jakab A., Bauer S., et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) // *IEEE Transactions on Medical Imaging*. 2015. Vol. 34, No. 10. P. 1993–2024. DOI: 10.1109/TMI.2014.2377694. <https://doi.org/10.1109/TMI.2014.2377694>
3. · Chen J., Lu Y., Yu Q., Luo X., Adeli E., Wang Y., Lu L., Yuille A.L., Zhou Y. TransUNet: Transformers make strong encoders for medical image segmentation // *arXiv preprint arXiv:2102.04306*. 2021. [arXiv:2102.04306](https://arxiv.org/abs/2102.04306)
4. · Hatamizadeh A., Tang Y., Nath V., Yang D., Myronenko A., Landman B., Roth H.R., Xu D. UNETR: Transformers for 3D medical image segmentation // *Proc. IEEE/CVF WACV*. 2022. P. 574–584. DOI: 10.1109/WACV51458.2022.00181. <https://doi.org/10.1109/WACV51458.2022.00181>
5. · Hatamizadeh A., Nath V., Tang Y., Yang D., Roth H.R., Xu D. Swin UNETR: Swin Transformers for semantic segmentation of brain tumors in MRI images // *Proc. Int. MICCAI Brainlesion Workshop*. 2022. P. 272–284. DOI: 10.1007/978-3-031-08999-2_22. https://doi.org/10.1007/978-3-031-08999-2_22
6. · Zhou H.Y., Guo J., Zhang Y., Yu L., Wang L., Yu Y. nnFormer: Volumetric medical image segmentation via a 3D Transformer // *IEEE Transactions on Image Processing*. 2023. Vol. 32. P. 4036–4045. DOI: 10.1109/TIP.2023.3293771. <https://doi.org/10.1109/TIP.2023.3293771>
7. · Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., Guo B. Swin Transformer: Hierarchical vision Transformer using shifted windows // *Proc. IEEE/CVF ICCV*. 2021. P. 10012–10022. DOI: 10.1109/ICCV48922.2021.00986. <https://doi.org/10.1109/ICCV48922.2021.00986>
8. · Baid U., Ghodasara S., Mohan S., et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification // *arXiv preprint arXiv:2107.02314*. 2021. [arXiv:2107.02314](https://arxiv.org/abs/2107.02314)
9. · Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation // *Proc. MICCAI*. 2015. P. 234–241. DOI: 10.1007/978-3-319-24574-4_28. https://doi.org/10.1007/978-3-319-24574-4_28
10. · Dosovitskiy A., Beyer L., Kolesnikov A., et al. An image is worth 16x16 words: Transformers for image recognition at scale // *Proc. ICLR*. 2021. [Official publication page](https://openreview.net/forum?id=cm333843h2)

Proofreader: Zokir ALIBEKOV

Layout and Designer: Oloviddin Sobir ugli

2026. № 9

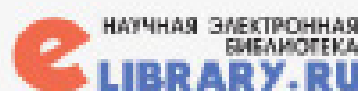
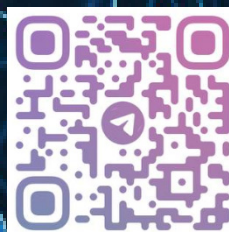
© When materials are reproduced, the INNOVATION SCIENCE AND TECHNOLOGY journal must be cited as the source. Authors are responsible for the accuracy of the information in materials and advertisements published in the journal. Editorial opinions may not always align with those of the authors. Submitted materials will not be returned to the editorial office.

To publish articles in this journal, you may submit articles, advertisements, stories, and other creative materials through the following links. Materials and advertisements are published on a paid basis.

You may subscribe to the journal at any time using the following details. Once subscribed, please send a screenshot or photo of your payment confirmation to our Telegram page @iqtisodiyot_77. Based on this, we will send the latest issue of the journal to your address each month.

“The journal “INNOVATION SCIENCE AND TECHNOLOGY” has been registered by the Agency for Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan from 09.10.2024 under the registration number №390637. License number: C-5669633. PNFL: 30407832680027

Our address: Tashkent city, Yunusobod district, 19th block,
House 17.



Acceptance of articles

Published every
monthly



Directions

Social, economic, political,
technological, scientific



Scopus || Scientific electronic journal specializing in Scopus

CERTIFICATE NUMBER: №390637

**ORDER NUMBER ACCORDING TO
THE LICENSE REGISTER: C-5669633**

CONTACT:



Contact us
+998 50 737 87 88



Telegram channel
t.me/scopus_IST2100



Journal official website
<https://ist-journal.uz/index.php/IST>