УДК 004.8

С. Л. Козлов; О. К. Колесницький, канд. техн. наук, проф.

ЕФЕКТИВНІ ДИФУЗІЙНІ МОДЕЛІ ДЛЯ SUPER-RESOLUTION ЗОБРАЖЕНЬ

Дифузійні моделі встановили нові стандарти перцептивної якості у SISR, проте їхнє багатокрокове висновування та великий розмір моделі ускладнюють практичне розгортання: моделі на основі Stable Diffusion потребують 50–200 кроків знешумлення, секундні затримки та мільярди параметрів. Цей огляд систематизує два взаємодоповнюючі напрямки: ефективне проєктування дифузійного процесу, що скорочує ітеративне семплювання від сотень кроків до кількох, та ущільнення моделей для розгортання з обмеженими ресурсами. Проаналізовано дванадцять моделей 2023–2025 років: вісім ефективних (ResShift, SinSR, OSEDiff, TSD-SR, AddSR, DoSSR, CCSR, InvSR) та чотири ущільнених (AdcSR, PassionSR, Edge-SD-SR, BiMaCoSR), та проведено порівняння їх за якістю (SSIM, LPIPS, CLIPIQA, MUSIQ) та ефективністю (параметри, MACs, час висновування) на тестових наборах DIV2K, RealSR і DRealSR. 3-поміж ефективних моделей ті, що побудовані на попередньо навчених text-to-image опорних моделях, дають приріст до +0,13 CLIPIQA порівняно з моделями навченими з нуля. Використання LR-зображення, як початкової точки зворотного процесу, забезпечує кращий баланс перцепція-спотворення, порівняно зі початком з гаусового шуму. Моделі-студенти можуть показати кращі результати, ніж відповідні моделі-вчителі, за умови донавчання на еталонних зображеннях. Текстові запити слугують допоміжним, а не обов'язковим сигналом. InvSR та CCSR виносять баланс перцепція-спотворення як runtime-параметр на єдиній навченій моделі. Ущільнення у 4-6 разів майже не впливає на якість. Понад 10-кратне ущільнення погіршує перцептивну якість, хоча точність відтворення зберігається. VAE-декодер домінує в обчисленнях та затримці на пристрої, що робить його першочерговою ціллю ущільнення. Водночас ущільнені дифузійні SR-моделі, все ще, значно більші за GAN-моделі, і оптимальний компроміс між розміром моделі та якістю результату залишається недослідженим.

Ключові слова: *image super-resolution, diffusion models, однокрокові дифузійй моделі, ущільнення дифузійних моделей, дистиляція знань, Stable Diffusion, глибоке навчання.*

Вступ

Задача super-resolution одного зображення (SISR, single image super-resolution) полягає у відновленні зображення високої роздільності (HR, high-resolution) з його відповідника низької роздільності (LR, low-resolution). SISR – некоректно поставлена обернена задача, що вимагає, як точності відтворення оригіналу, так і, перцептивного реалізму. Ці цілі за своєю природою протилежні: Blau Y. та ін. [1] довели, що жоден алгоритм не може одночасно мінімізувати спотворення та максимізувати перцептивну якість, тож кожна модель SR (super-resolution) неминуче шукає компроміс між ними. Прикладна (real-word) задача SR додатково ускладнена тим, що спотворення невідоме й може поєднувати розмиття, шум, стиснення та зменшення роздільності у довільному порядку – на відміну від класичного SR, який припускає фіксоване ядро спотворення, наприклад бікубічне.

Методи SR на основі глибокого навчання пройшли шлях від CNN-регресії (ESPCN [2], RCAN [3]) та мереж із самоувагою (SwinIR [4]), орієнтованих на точність відтворення, до GAN-підходу (Real-ESRGAN [5]) із фокусом на реалістичність сприйняття. Дифузійні моделі знешумлення (DM, denoising Diffusion Models) встановили новий рівень перцептивної якості в задачі SR, проте ціною значних обчислень: першопрохідна SR3 [6] потребувала 1000 кроків знешумлення (~30с на зображення). StableSR [8], SeeSR [9], DiffBIR [10] залучили генеративний потенціал Stable Diffusion (SD) із мільярдами параметрів до задачі SR, досягнувши вищої перцептивної якості та скоротивши кількість кроків до 50–200, а затримку – до 4–10 с. Втім, вимоги до часу

висновування (inference) та пам'яті залишаються доволі високими для широкого практичного застосування.

Наявні огляди широко охоплюють дифузійні методи реставрації зображень. Не та ін. [11] розглядають DM у понад двадцяти низькорівневих задачах комп'ютерного зору, без зосередження на аналізі ефективності SR; Li та ін. [12] оглядають дифузійну реставрацію та відновлення зображень, акцентуючи на парадигмах навчання; Moser та ін. [13] пропонують SR-орієнтовану таксономію, але передують хвилі однокрокових та ущільнених моделей 2024 – 2025 років. Жоден із них не розглядає ефективність, як центральне питання, і не порівнює її разом з якістю на спільних тестових наборах.

Це коротке оглядове дослідження заповнює зазначену прогалину порівняльним аналізом дванадцяти ефективних моделей прикладної SR: вісім із них адаптують дифузійний процес до одно- чи малокрокового висновування, чотири – ущільнюють архітектуру для розгортання.

Передумови

DM генерують зображення, навчаючись обертати заздалегідь відомий процес зашумлення. Ймовірнісна дифузійна модель знешумлення (DDPM, denoising diffusion probabilistic model) [14] визначає *прямий процес*, що поступово додає гаусів шум до чистого зображення x_0 впродовж T часових кроків, створюючи версії x_t зі все більшим рівнем шуму:

$$q(x_t|x_0) = \mathcal{N}\left(x_t;\ \sqrt{\bar{\alpha}_t}\, x_0,\ (1-\bar{\alpha}_t)\, \boldsymbol{I}\right), \quad (1)$$

де $\bar{\alpha}_t$ – кумулятивний розклад шуму, що контролює, скільки сигналу залишається на кроці t. Мережа UNet $\epsilon_\theta(\mathrm{x_t}, \mathrm{t})$ навчається передбачати доданий шум, що уможливлює *зворотний процес* – покрокове знешумлення від чистого шуму назад до чистого зображення. Більша кількість кроків дозволяє досягти вищої якості, але пропорційно збільшує затримку: SR3 [6] потребувала 1000 кроків.

Латентні дифузійні моделі (LDM, Latent Diffusion Models) [7] – основа сімейства SD – переводять дифузійний процес у латентний простір за допомогою варіаційного автокодера (VAE, Variational Autoencoder) із просторовим стисненням у 4 рази. Такий перехід прискорив навчання та висновування приблизно у 3 рази порівняно з піксельним простором.

Більшість сучасних моделей SR застосовують SD як опорну модель (backbone). SD поєднує UNet, VAE та текстовий кодер (SD 2.1 ~1,3 млрд параметрів загалом). SD-Turbo [15] завдяки змагальній дифузійній дистиляції (ADD, Adversarial Diffusion Distillation) зводить висновування до одного кроку. SD 3 [16] замінює UNet на Diffusion Transformer (DiT) та переходить до узгодження потоків (FM, flow matching) (~2,5 млрд параметрів).

Для розв'язання задачі SR за допомогою DM сформувались два основні підходи. Перший – навчання спеціалізованих дифузійних моделей з нуля. Другий – використання попередньо навчених T2I-моделей із замороженими або мінімально адаптованими вагами, де SR-обумовлення впроваджується через ControlNet [17] або донавчання через LoRA (Low-Rank Adaptation). Другий підхід жертвує компактністю заради вищої перцептивної якості.

Ефективні дифузійні моделі

Таблиця 1 підсумовує вісім розглянутих ефективних моделей разом із чотирма базовими моделями для порівняння. Значення метрик взято з оригінальних статей, уніфікованої переоцінки Sun та ін. [18] та аналітичних оцінок (якщо зазначено). Обговорення організовано навколо двох ліній: моделі, з оригінальною архітектурою, навчені з нуля, та моделі, побудовані на основі уже наявних опорних T2I-моделей.

Ефективна дифузія без попередньонавчених моделей

ResShift [19] моделює дифузійний процес через залишковий зсув:

$$q(x_t \mid x_0, y) = \mathcal{N}(x_t;\ x_0 + \eta_t(y - x_0),\ \kappa^2\eta_t \boldsymbol{I}), \quad (2)$$

де x_0 – HR-зображення, y – LR-зображення, η_t – монотонно зростаюча вага ($\eta_0 \to 0, \eta_T \to 1$), яка контролює величину зсуву, κ – масштабує дисперсію шуму. Це зміщує дифузійну траєкторію, від повного переходу шум-зображення до значно коротшого LR-HR переходу, тому 15 кроків достатньо для конкурентних результатів. UNet працює у латентному просторі VQGAN із блоками Swin Transformer; 173,9M параметрів, 0,76с, найвищий SSIM з-поміж розглянутих моделей.

SinSR [20] – однокрокова модель, отримана методом дистиляції знань з моделі ResShift. Дистиляційна втрата [21] узгоджує результат моделі-студента з виходом моделі-вчителя, проте потенційна якість моделі-студента обмежена якістю моделі-вчителя. З метою подолати це обмеження, студента навчають інвертувати власний результат – за згенерованим зображенням відновлювати вхідний шум. Наступним етапом інверсію застосовують до еталонних зображень (GT, ground truth): GT перетворюють у відповідний шум, з якого студент має відтворити оригінал – і таким чином навчається безпосередньо на еталонних даних, а не лише на передбаченнях вчителя. З 118,59M навчальних параметрів та ~0,13с SinSR покращує результати ResShift за перцептивними метриками (CLIPIQA 0,6887 проти 0,5958 на RealSR).

Ефективна дифузія на основі попередньонавчених опорних моделей

OSEDiff [22] – перша однокрокова модель SR на основі SD, яка починає зворотний процес із латентного представлення LR-зображення замість гаусового шуму. Модель адаптує Variational score distillation (VSD) [23] до латентного простору, де зафіксована SD виступає моделлю-вчителем, а LoRA-адаптована версія – моделлю-студентом. Degradation-aware prompt extractor (DAPE) із SeeSR [9] генерує текстові запити (prompt) для LR-зображень і подає їх як моделі-студенту (під час навчання та висновування), так і моделі-вчителю. З LoRA 4-го рангу навчальними є лише 8,5M із 1775M параметрів; висновування займає 0,11 с. Експерименти на DRealSR показали, що відмова від текстових запитів покращує точність відтворення (SSIM 0,791 проти 0,7835), але знижує перцептивну якість (CLIPIQA 0,6599 проти 0,6963).

TSD-SR [24] вдосконалює VSD у двох напрямках. Target Score Matching (TSM) використовує GT-латенти як опорну точку для моделі-вчителя, що забезпечує надійніші градієнти порівняно з VSD. Distribution-Aware Sampling Module (DASM) зосереджує навчання на часових кроках, найкритичніших для відновлення текстурних деталей, замість рівномірного семплювання. Під час дистиляції модель-вчитель керується текстовими запитами, отриманими з GT-зображень, тоді як модель-студент використовує константне вбудування, як під час дистиляції, так і під час висновування. Латент LR-зображення є початковою точкою зворотного процесу, як і в OSEDiff. Побудована на SD 3 з LoRA 64-го рангу, TSD-SR досягає найкращого LPIPS на всіх тестових наборах (0,2673/0,2743/0,2967) та найкращого MUSIQ на DIV2K (71,69) за ~0,136с. Порівняння на SD 2.1 підтверджує перевагу TSM над VSD.

AddSR [25] дистилює SeeSR у малокрокову модель за допомогою ADD, де змагальна втрата (дискримінатор) забезпечує перцептивний реалізм результатів моделі-студента, а дистиляційна втрата – узгодженість з результатами моделі-вчителя. Оскільки фіксований баланс цих двох втрат в оригінальному ADD спричиняє розмиття при малій кількості кроків і галюцинації при великій, AddSR вводить надбудову Timestep-adaptive ADD (TA-ADD), яка динамічно зміщує цей баланс залежно від часового кроку. Текстові запити генеруються на основі RAM [26] тегів та кодуються CLIP [27] і керують процесом, як під час дистиляції, так і під час висновування. Обумовлення процесу відбувається через ControlNet: на першому кроці – LR-зображенням, а на наступних – передбаченим HR-зображенням з попереднього кроку завдяки механізму Prediction-based self-

refinement (PSR). За 4 кроки, 0,81с та при 2510M параметрів AddSR досягає найвищого CLIPIQA (0,7794/0,7215/0,7381), при найнижчому SSIM (0,5651/0,6336/0,7036).

DoSSR [28] формулює прямий процес, як зміщення домену (domain shift), де середнє розподілу лінійно інтерполює між LR та HR, тоді як, розклад шуму відповідає стандартному DDPM. На відміну від ResShift, чий розклад шуму несумісний з попередньо навченою SD, DoSSR зберігає розклад оригінального SD, що дає змогу безпосереднього донавчати модель замість дистиляції. LR-обумовлення слідує архітектурі ControlNet з DiffBIR; текстове обумовлення не використовується. За 5 кроків, ~0,55 с та при 1716,6M параметрів досягає CLIPIQA 0,7014/0,7025/0,6776.

CCSR [18] розв'язує компроміс між відтворюваністю та якістю: стохастичне семплювання дає багаті деталі, але непостійні результати між запусками. Non-Uniform Timestep Sampling (NUTS) використовує лише кілька кроків знешумлення для відновлення структури, обриваючи процес до того, як генерація деталей внесе нестабільність. Далі, текстури детерміновано покращує DeFT – VAE-декодер, донавчений за допомогою GAN-втрати. Текстове обумовлення не використовується. З розміром у 1650M параметрів, 2 кроками та ~0,17 с CCSR досягає найвищого MUSIQ на DRealSR (68,49) і пропонує однокрокове або двокрокове висновування в межах однієї моделі.

InvSR [29] обирає найрадикальніший підхід: уся опорна модель SD-Turbo залишається зафіксованою. Навчається невеликий зовнішній предиктор шуму (33,84M), який додає відповідний шум до закодованого VAE латенту LR-зображення, створюючи слабко зашумлений стан. SD-Turbo потім знешумлює цей стан безпосередньо за один крок. Без ControlNet, LoRA чи текстових запитів для конкретного зображення – лише константний текстовий опис. При ~0,117 с висновування InvSR показує кращі результати за OSEDiff за всіма 7 метриками на ImageNet-Test, як повідомляють автори, і підтримує 1–5 кроків без перенавчання.

Таблиця 1

Порівняння ефективних дифузійних моделей для вирішення задачі SR

Назва моделі [дата публ.]	**Базова модель**	**К-сть параметрів (M)** [к-сть параметрів, що навчаються]	**Час виснов. (A100, 128→512)**	**Кроки**	**Тестові набори**			**Метрика**
					DIV2K[30]	**RealSR**[31]	**DRealSR**[32]	
R-ESRGAN[5] [07.21]	—	16,7	0,065s	1	0,6372†	0,7616†	0,8053†	▲ SSIM[33]
					0,3124†	0,2727†	0,2847†	▼ LPIPS[34]
					0,5219†	0,4449†	0,4422†	▲ CLIPIQA[35]
					60,92†	60,18†	54,18†	▲ MUSIQ[36]
ResShift[19] [07.23]	—	173,9 [118,59]	0,76s†	15	**0,6175**†	**0,7411**†	0,7632†	▲ SSIM
					0,3374†	0,3489†	0,4073†	▼ LPIPS
					0,6089†	0,5450†	0,5259†	▲ CLIPIQA
					60,92†	58,10†	49,86†	▲ MUSIQ
SinSR[20] [11.23]	ResShift	173,9 [118,59]	0,13s†	1	0,6012†	**0,7354**†	0,7495†	▲ SSIM
					0,3262†	0,3212†	0,3741†	▼ LPIPS
					0,6499†	0,6204†	0,6367†	▲ CLIPIQA
					62,80†	60,41†	55,34†	▲ MUSIQ
StableSR[8] [05.23]	SD 2.1	1410‡ [150‡]	10,03s†	200	0,5726	0,7080	0,7536	▲ SSIM
					0,3114	0,3002	0,3284	▼ LPIPS
					0,6771	0,6234	0,6357	▲ CLIPIQA
					65,92	65,88	58,51	▲ MUSIQ
DiffBIR[10] [08.23]	SD 2.1	1717‡ [380‡]	2,72s†	50	0,5653†	0,6673†	0,6660†	▲ SSIM
					0,3541†	0,3567†	0,4446†	▼ LPIPS
					0,6652†	0,6412†	0,6292†	▲ CLIPIQA
					65,66†	64,66†	60,68†	▲ MUSIQ

Продовження таблиці 1

Назва моделі [дата публ.]	Базова модель	К-сть параметрів (M) [к-сть параметрів, що навчаються]	Час виснов. (A100, 128→512)	Кроки	Тестові набори			Метрика
					DIV2K[30]	RealSR[31]	DRealSR[32]	
SeeSR[9] [11.23]	SD 2.1	2524‡ [749,9‡]	4,30s†	50	0,5386	0,7216	0,7691	▲ SSIM
					0,3843	0,3009	0,3189	▼ LPIPS
					0,6946	0,6612	0,6804	▲ CLIPIQA
					68,33	69,77	64,93	▲ MUSIQ,
CCSR[18] [12.23]	SD 2.1	1650	0,17s	2	**0,6130**	0,7335	**0,7724**	▲ SSIM
					0,3152	0,2941	0,3397	▼ LPIPS
					0,7000	0,6561	0,6695	▲ CLIPIQA
					71,65	71,17	**68,49**	▲ MUSIQ
AddSR[25] [04.24]	SeeSR	2510§	0,81s	4	0,5651	0,6336	0,7036	▲ SSIM
					0,2812	0,3742	0,3866	▼ LPIPS
					0,7794	**0,7215**	**0,7381**	▲ CLIPIQA
					71,43	**72,25**	**68,16**	▲ MUSIQ
OSEDiff[22] [06.24]	SD 2.1	1775 [8,5]	0,11s	1	0,6108	0,7341	**0,7835**	▲ SSIM
					0,2941	0,2921	**0,2968**	▼ LPIPS
					0,6683	0,6693	0,6963	▲ CLIPIQA
					67,97	69,09	64,65	▲ MUSIQ
DoSSR[28] [09.24]	SD	1716,6	~0,55s*	5	0,6073	0,6839	0,7298	▲ SSIM
					0,3371	0,3374	0,3689	▼ LPIPS
					0,7014	0,7025	0,6776	▲ CLIPIQA
					66,54	69,42	64,40	▲ MUSIQ
TSD-SR[24] [11.24]	SD 3	—	0,136s	1	0,5808	0,7172	0,7559	▲ SSIM
					0,2673	**0,2743**	**0,2967**	▼ LPIPS
					0,7416	**0,7160**	**0,7344**	▲ CLIPIQA
					71,69	**71,19**	66,62	▲ MUSIQ
InvSR[29] [12.24]	SD-Turbo	~1327* [33,84]	0,117s	1	—	0,7262	—	▲ SSIM
					—	**0,2872**	—	▼ LPIPS
					—	0,6918	—	▲ CLIPIQA
					—	67,46	—	▲ MUSIQ

best value – найкраще значення у наборі (набір даних, метрика); базові моделі для порівняння не враховуються
second-best – друге найкраще значення у наборі (набір даних, метрика); базові моделі для порівняння не враховуються
† – Sun та ін. [18] Табл. IV/VI
‡ – Wu та ін. [22] Табл. 2
§ – Sun та ін. [18] Табл. VII
¶ – значення ефективності (час висновування), виміряне на A100 40G, 128→512 ×4 SR – Yue та ін. [29] Табл. IV (дод.)
* – оцінено/конвертовано: час DoSSR V100→A100 через міжмодельне співвідношення 0,531; загальний розмір InvSR через проксі S3Diff – Chen та ін. [25] Табл. 1

Порівняльний аналіз та тенденції

Вплив T2I опорної моделі. Моделі, навчені з нуля (ResShift, SinSR), показують високий SSIM, але низькі показники перцептивної якості; моделі на основі T2I забезпечують вищі значення – до +0,13 CLIPIQA (SinSR 0,6204 проти AddSR 0,7215 на RealSR). InvSR, маючи лише 33,84M навчених параметрів на зафіксованій опорній моделі, досягає CLIPIQA 0,6918 на RealSR – що свідчить: високу перцептивну якість забезпечує саме попередньо навчена T2I опорна модель, а не обсяг адаптації.

Інтеграція LR. Спосіб введення LR-зображення у дифузійний процес визначає позицію на осі перцепція-спотворення. Конкатенація каналів (ResShift, SinSR) тісно зв'язує LR з UNet, забезпечуючи найвищий SSIM, але найслабші перцептивні показники. Обумовлення через ControlNet (AddSR, CCSR, DoSSR) обмежує зв'язок LR – UNet, зберігаючи повну генеративну незалежність SD: AddSR досягає найвищого CLIPIQA (0,7215), але найнижчого SSIM (0,6336) на RealSR. LR-латент, як початкова точка зворотнього процесу (OSEDiff, TSD-SR), забезпечує

найкращий баланс перцепція-спотворення – UNet розглядає деградації LR, як шум, що потрібно видалити, досягаючи найкращого LPIPS (TSD-SR 0,2743 на RealSR). InvSR доводить це до крайності: зафіксований UNet ніколи не отримує LR безпосередньо, максимізуючи використання знань опорної моделі. Коротко: тісніший зв'язок LR–UNet покращує точність відтворення; слабший – перцептивну якість.

Дистиляція. Підходи варіюються від тісного зв'язку з моделлю-вчителем до повної незалежності: VSD (OSEDiff) зіставляє передбачення шуму опорної та LoRA-адаптованої моделі на зашумлених латентах, даючи слабкий навчальний сигнал; TSM (TSD-SR) порівнює передбачення моделі-вчителя на згенерованих та зашумлених GT-латентах, даючи якісніші сигнали; ADD (AddSR) натомість звертається до дискримінатора; InvSR та CCSR взагалі обійшлись без дистиляції. Кілька моделей-студентів досягли вищих показників якості за своїх моделей-вчителів (SinSR, TSD-SR, AddSR) –переважно завдяки залученню GT-зображень або у функціях втрат, або для обумовлення моделі-вчителя.

Текстове керування. Експерименти OSEDiff показують, що відмова від текстових запитів підвищує точність відтворення (SSIM +0,008), але знижує перцептивну якість (CLIPIQA −0,036 на DRealSR). Текст не є обов'язковим: CCSR та DoSSR не використовують його взагалі, причому CCSR досягає найвищого MUSIQ на DRealSR (68,49). InvSR використовує фіксований загальний текстовий запит. Текстове обумовлення покращує текстури в межах T2I-архітектур, але LR-зображення залишається основним сигналом для SR.

Версії SD опорної моделі. SD 2.1 лежить в основі OSEDiff, CCSR, DoSSR та AddSR (через SeeSR). TSD-SR переходить на SD 3 (DiT +FM), демонструючи найкращі LPIPS та MUSIQ. InvSR використовує SD-Turbo; перехід на FM формується, як тенденція, але поки зарано говорити про усталений тренд.

Швидкість проти якості. Однокрокові моделі (0,11с–0,136 с) формують конкурентний рівень; AddSR із 4 кроками (~0,81 с) досягає найвищих перцептивних результатів при 6–8х більшій затримці, проте, навіть, 0,11 с залишається надто повільним для задач реального часу, пакетної обробки або розгортання на кінцевих пристроях – підкреслюючи потребу в ущільненні.

Варто виділити шість тенденцій: висновування за кілька кроків (1–5) стало стандартом для прикладної SR; тренування з нуля поступилося попередньо навченим T2I опорним моделям, що призвело до значного зростання вимог до пам'яті; адаптація спростилася від важкого донавчання через LoRA до зафіксованих опорних моделей; початкові точки зворотного процесу еволюціонували від зашумленого LR-латенту до вивченої інверсії; FM починає з'являтися, як альтернатива дискретно-кроковій дифузії; контроль компромісу перцепція-спотворення стає явним, водночас InvSR та CCSR виносять його як runtime-параметр.

Ущільнені дифузійні моделі

Ефективне висновування скорочує кількість кроків від сотень до кількох, проте розміри моделей залишаються незмінними – від сотень мільйонів до мільярдів параметрів. Три ортогональні стратегії ущільнення спрямовані на цей розрив. Таблиця 2 порівнює чотири ущільнені моделі разом із SinSR та OSEDiff, як неущільненими орієнтирами. Значення взяті з оригінальних статей; перехресно цитовані та оцінені записи позначені.

AdcSR [37] структурно ущільнює OSEDiff. VAE-кодер замінено на PixelUnshuffle [2], який зберігає всю вхідну інформацію; експерименти підтверджують підвищення точності відтворення (PSNR +0,13 дБ) та перцептивної якості (DISTS [38] −0,004) на DRealSR. Текстовий кодер, перехресна увага та вкладення часового кроку видалені з незначним впливом. Канали UNet обрізано на 25 %, VAE-декодер – на 50 %. Для відновлення якості модель-студент навчалась з ознак вчителя через L1-втрати та з GT зображень через дискримінатор. Модель на 456M параметрів (−74 %) покращує безеталонні метрики, порівняно з OSEDiff на DRealSR (CLIPIQA

+0,009, MUSIQ +1,61), з помірними поступками у точності відтворення (SSIM −0,011, LPIPS +0,008).

PassionSR [39] застосовує квантизацію після навчання (PTQ, post-training quantization) з точністю W8A8/W6A6 до OSEDiff після заміни DAPE та CLIP-кодер на константне вкладання (−27 % параметрів). Оскільки VAE домінує в обчисленнях (>80 % MAC-операцій), спільна квантизація UNet-VAE є критичною. W8A8 дає 238M (−82 %), ~530G MAC-операцій, втрати <1 % та найкращий CLIPIQA з-поміж ущільнених моделей (0,6939/0,6912/0,7554). W6A6 (178M) потрапляє на поріг якості: MUSIQ падає з 65,88 до 44,43 на RealSR.

Edge-SD-SR [40] ущільнює SD 1.5 для мобільного розгортання: UNet зменшується з 860M до 158M параметрів, автокодер – з 83M до 14M. Ключова ідея –двонапрямлене обумовлення: LR-зображення впливає не лише на зворотний процес (через конкатенацію), а й на прямий, де статистики LR-зображення (середнє та дисперсія) вбудовуються безпосередньо у розподіл шуму. При ~169M параметрів та 142 GFLOPs це єдина модель, протестована на мобільному пристрої: 38 мс на Samsung S24 NPU для 128→512 із найкращим LPIPS серед ущільнених моделей (0,2490/0,2780/0,2920).

BiMaCoSR [41] застосовує екстремальну 1-бітну бінаризацію до SinSR, доводячи лінію ResShift/SinSR до граничного ущільнення. Наївна бінаризація спричиняє колапс через втрату інформації та невідповідність розподілів. Три окремі гілки запобігають цьому: BMB (бінаризована, XNOR/bit-count), LRMB (низькорангова, ранг 8, SVD-ініціалізована) для низькочастотного вмісту та SMB (розріджена) для поглинання викидів, що спільно зменшують помилку квантизації з 1,1275 до 0,1855. При 4,98M параметрів та 1,83G FLOPs (23,8x ущільнення) точність відтворення, згідно з даними статті (вимірювання при 64→256), зберігається (SSIM 0,7547/0,7698/0,8393), але перцептивна якість погіршується (ΔMUSIQ −11,24/−8,86/−1,43 порівняно з SinSR) на DIV2K/RealSR/DRealSR.

Таблиця 2

Порівняння ущільнених дифузійних моделей для вирішення задачі SR

Назва моделі [дата публ.]	Базова модель	К-сть параметрів (M) [к-сть параметрів, що навчаються]	MACs (G)	Тестові набори			Метрика
				DIV2K[30]	RealSR[31]	DRealSR[32]	
SinSR[20] [11.23]	ResShift	173,9 [118,59]	2,649‡	0,6012†	0,7354†	0,7495†	▲ SSIM[33]
				0,3262†	0,3212†	0,3741†	▼ LPIPS[34]
				0,6499†	0,6204†	0,6367†	▲ CLIPIQA[35]
				62,80†	60,41†	55,34†	▲ MUSIQ[36]
OSEDiff[22] [06.24]	SD 2.1-base	1775 [8,5]	2265	0,6108	0,7341	0,7835	▲ SSIM
				0,2941	0,2921	0,2968	▼ LPIPS
				0,6683	0,6693	0,6963	▲ CLIPIQA
				67,97	69,09	64,65	▲ MUSIQ
AdcSR[37] [11.24]	OSEDiff (SD 2.1)	456	496	—	—	0,7726	▲ SSIM
				—	**0,2885**	**0,3046**	▼ LPIPS
				—	**0,6731**	**0,7049**	▲ CLIPIQA
				—	—	**66,26**	▲ MUSIQ
PassionSR[39] W8A8 [11.24]	OSEDiff (SD 2.1)	238	530*	**0,7199**	**0,7499**	**0,8146**	▲ SSIM
				0,2496	0,3140	0,3422	▼ LPIPS
				0,6939	**0,6912**	**0,7554**	▲ CLIPIQA
				67,92	**65,88**	33,56	▲ MUSIQ
Edge-SD-SR[40] [12.24]	SD 1.5	169	71*	0,6170	—	—	▲ SSIM
				0,2490	**0,2780**	**0,2920**	▼ LPIPS
				—	—	—	▲ CLIPIQA
				69,58	**65,20**	**55,66**	▲ MUSIQ

Продовження таблиці 2

Назва моделі [дата публ.]	Базова модель	К-сть параметрів (M) [к-сть параметрів, що навчаються]	MACs (G)	Тестові набори			Метрика
				DIV2K[30]	RealSR[31]	DRealSR[32]	
BiMaCoSR[41] [02.25]	SinSR	4,98	0,915*◇	**0,7547◇**	**0,7698◇**	**0,8393◇**	▲ SSIM
				0,2999◇	0,3375◇	0,3400◇	▼ LPIPS
				0,5176◇♭	0,4800◇♭	0,4867◇♭	▲ CLIPIQA
				53,38◇	49,01◇	29,38◇	▲ MUSIQ

best value – найкраще значення у наборі (набір даних, метрика); базові моделі для порівняння не враховуються
second-best – друге найкраще значення у наборі (набір даних, метрика); базові моделі для порівняння не враховуються
† – Sun та ін. [18] Табл. VI/VII
‡ – Wu та ін. [OSEDiff] Tab.2
* – оцінено/конвертовано: FLOPs→MACs через MACs ≈ FLOPs/2
◇ – значення при нестандартному розмірі 64→256
♭ – метрика CLIP-IQA+ (Wang та ін. 2023a), не CLIPIQA; виключено з рейтингу

Порівняльний аналіз та тенденції

Чутливість компонентів до ущільнення. Модуль текстового обумовлення виявився зайвим (видалений у всіх моделях). VAE-кодер є замінним: експерименти AdcSR підтверджують, що PixelUnshuffle [2] покращує, як точність відтворення, так і перцептивну якість, уникаючи втратного кодування у латентний простір. VAE-декодер є одночасно найважчим компонентом (>80 % MACs, 59 % затримки на пристрої) і найскладнішим для ущільнення без втрати перцептивної якості. UNet витримує обрізку до 75 % каналів (AdcSR) та бінаризацію (BiMaCoSR).

Компроміс ущільнення–якість. За помірного ущільнення (4–6x) втрата якості є незначною: AdcSR (3,9x) покращує безеталонні показники (CLIPIQA +0,009, MUSIQ +1,61 на DRealSR) з помірними поступками у точності відтворення; PassionSR W8A8 (5,5x) втрачає < %; Edge-SD-SR (~5,6x) досягає найкращого LPIPS. При ущільненні понад 10x точність відтворення та перцептивна якість розходяться: BiMaCoSR (23,8x) зберігає високий SSIM, але втрачає 8,86 MUSIQ; PassionSR W6A6 демонструє те саме – MUSIQ колапсує (44,43 проти 65,88 на RealSR). Три моделі на основі SD (169 – 456M) досягають MUSIQ 55,66 – 66,26 на DRealSR; BiMaCoSR, побудована без SD, обмежується 29,38. Визначення порогу ущільнення, за яким перцептивна якість починає різко падати, є ключовим для встановлення оптимального розміру моделі для розгортання.

Мобільне розгортання. Edge-SD-SR є єдиним підтвердженням концепції на пристрої. 0,03 с AdcSR на A100 свідчить про мобільний потенціал, але не має даних NPU; цілочисельна арифметика PassionSR придатна для NPU, але не має показників затримки. Систематичне розгортання на кінцевих пристроях залишається відкритою проблемою.

Варто виділити п'ять тенденцій: три підходи до ущільнення спрямовані на незалежні типи надмірності та можуть комбінуватися; модуль текстового обумовлення виявився зайвим, і видалений в усіх ущільнених моделях; VAE-декодер є домінуючим вузьким місцем, що робить декодер-орієнтоване ущільнення найпріоритетнішою ціллю; галузь рухається від пост-фактум ущільнення до спільного проєктування (Edge-SD-SR); точність відтворення витримує ущільнення краще, ніж перцептивна якість; збереження здатності до синтезу текстур при ущільненні потребує спеціальних методів.

Висновки

Проаналізовано дванадцять дифузійних SR-моделей, опублікованих між 2023 та 2025 роками – вісім спрямованих на ефективне проєктування дифузійного процесу та чотири на ущільнення моделей – з кількісним порівнянням їхньої якості, швидкості та розміру на спільних тестових наборах.

Попередньо навчена T2I опорна модель має найбільший вплив на перцептивну якість – перевага сягає +0,13 CLIPIQA порівняно з моделями, навченими з нуля, хоча ціною ~10x більшого розміру моделі. InvSR показує мінімалістичний шлях: зафіксована опорна модель SD-Turbo з 33,84M навчальних параметрів дає конкурентну якість, а отже, за достатнього обсягу пам'яті, опорні моделі дедалі більше використовуватимуться як є, без модифікацій.

Моделі, що використовують LR як початкову точку дифузії, стабільно досягають кращого балансу перцепція-спотворення, ніж подання LR через ControlNet при старті з гаусового шуму. Текстове обумовлення надає додаткове керування текстурами в T2I-архітектурах, але не є необхідним – CCSR та DoSSR досягають найвищих перцептивних показників без керування текстом, що підтверджує: LR-зображення є основним сигналом обумовлення для SR.

Виходячи за межі рішень, закладених на етапі дизайну, InvSR та CCSR роблять наступний крок – виносять баланс перцепція-спотворення як runtime-параметр: єдина навчена модель дозволяє користувачу керувати балансом на етапі висновування.

Підходи до дистиляції утворюють спектр від повної залежності від вчителя (VSD) через порівняння з GT-латентами (TSM) та керування через дискримінатор (ADD) до повної відсутності вчителя (InvSR, CCSR). Кілька моделей-студентів досягли вищої якості за своїх вчителів, долаючи верхню межу дистиляції знань, завдяки функціям втрат із залученням GT-зображень або GT-обумовленим вчителям.

Ущільнення до 4-6x (AdcSR, PassionSR W8A8) спричиняє незначну втрату якості. Понад 10x точність відтворення зберігається, але перцептивна якість деградує – синтез текстур вимагає вищої точності, ніж піксельна регресія. Цей поріг оцінений емпірично, та жодна з робіт явно його не досліджує. VAE-декодер домінує в обчисленнях та затримці на пристрої (59 % [Edge-SD-SR]), що робить його першочерговою ціллю ущільнення.

Можемо виокремити кілька ширших тенденцій. По-перше, висока перцептивна якість вимагає або попередньо навченої опорної моделі, або дорогого навчання з нуля – простішого шляху, на разі, не знайдено. По-друге, ефективне висновування вже добре усталене: однокрокові моделі із зафіксованою опорною моделлю та без текстового обумовлення надають практичний спосіб залучити попередньонавчені T2I моделі для прикладної SR. По-третє, ущільнення дифузійних SR-моделей залишається на ранніх етапах розвитку – помірне ущільнення є безкоштовним, але навіть найменші моделі на основі SD (169–456M) у 10-27 разів більші за GAN-моделі, такі як Real-ESRGAN (16,7M), тоді як екстремальна бінаризація (BiMaCoSR, 4,98M) досягає малого розміру лише ціною перцептивної якості. Оптимальний компроміс між розміром моделі та якістю результату залишається недослідженим.

СПИСОК ЛІТЕРАТУРИ

1. Blau Y., Michaeli T. The Perception-Distortion Tradeoff. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 18–23 June 2018. 2018. URL: https://doi.org/10.1109/cvpr.2018.00652 (Last accessed: 22.02.2026).

2. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network / W. Shi et al. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016. 2016. URL: https://doi.org/10.1109/cvpr.2016.207 (Last accessed: 22.02.2026).

3. Image Super-Resolution Using Very Deep Residual Channel Attention Networks / Y. Zhang et al. *Computer Vision – ECCV 2018*. Cham, 2018. P. 294–310. URL: https://doi.org/10.1007/978-3-030-01234-2_18 (Last accessed: 22.02.2026).

4. SwinIR: Image Restoration Using Swin Transformer / J. Liang et al. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, 11–17 October 2021. 2021. URL: https://doi.org/10.1109/iccvw54120.2021.00210 (Last accessed: 22.02.2026).

5. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data / Wang X., Xie L., Dong C., Shan Y. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, 11–17 October 2021. 2021. URL: https://doi.org/10.1109/iccvw54120.2021.00217 (Last accessed: 22.02.2026).

6. Image Super-Resolution via Iterative Refinement / C. Saharia et al. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2023. Vol. 45, no. 4. P. 4713–4726. URL: https://doi.org/10.1109/tpami.2022.3204461 (Last accessed: 22.02.2026).

7. High-Resolution Image Synthesis with Latent Diffusion Models / R. Rombach et al. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 18–24 June 2022. 2022. URL: https://doi.org/10.1109/cvpr52688.2022.01042 (Last accessed: 22.02.2026).

8. Exploiting Diffusion Prior for Real-World Image Super-Resolution / J. Wang et al. *International Journal of Computer Vision. 2024.* Vol. 132, no. 12. P. 5929–5949. URL: https://doi.org/10.1007/s11263-024-02168-7 (Last accessed: 22.02.2026).

9. SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution / R. Wu et al. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2024. P. 25456–25467. URL: https://doi.org/10.1109/CVPR52733.2024.02405 (Last accessed: 22.02.2026).

10. DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior / X. Lin et al. *Computer Vision – ECCV 2024: 18th European Conference*, Milan, Italy, September 29–October 4, 2024, Proceedings, Part LIX, Milan, Italy. 2024. P. 430–448. URL: https://dl.acm.org/doi/10.1007/978-3-031-73202-7_25 (Last accessed: 22.02.2026).

11. Diffusion Models in Low-Level Vision: A Survey / C. He et al. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2025. P. 1–20. URL: https://doi.org/10.1109/tpami.2025.3545047 (Last accessed: 22.02.2026).

12. Diffusion Models for Image Restoration and Enhancement: A Comprehensive Survey / X. Li et al. International Journal of Computer Vision. 2025. Vol. 133, no. 11. P. 8078–8108. URL: https://doi.org/10.1007/s11263-025-02570-9 (Last accessed: 22.02.2026).

13. Diffusion Models, Image Super-Resolution And Everything: A Survey / B. B. Moser et al. IEEE Transactions on Neural Networks and Learning Systems. 2025. Vol. 36, no. 7. P. 11793–11813. URL: https://doi.org/10.1109/tnnls.2024.3476671 (Last accessed: 22.02.2026).

14. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models. *Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver,* BC, Canada. 2020. URL: https://dl.acm.org/doi/abs/10.5555/3495724.3496298 (Last accessed: 22.02.2026).

15. Adversarial Diffusion Distillation / Sauer A., Lorenz D., Blattmann A., Rombach R. *Computer Vision – ECCV 2024: 18th European Conference*, Milan, Italy, September 29–October 4, 2024, Proceedings, Part LXXXVI, Milan, Italy. 2024. P. 87–103. URL: https://dl.acm.org/doi/10.1007/978-3-031-73016-0_6 (Last accessed: 22.02.2026).

16. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis / P. Esser et al. *Proceedings of the 41st International Conference on Machine Learning*, Vienna, Austria. 2024. URL: https://dl.acm.org/doi/10.5555/3692070.3692573 (Last accessed: 22.02.2026).

17. Zhang L., Rao A., Agrawala M. Adding Conditional Control to Text-to-Image Diffusion Models. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2023. P. 3813–3824. URL: https://doi.org/10.1109/ICCV51070.2023.00355 (Last accessed: 22.02.2026).

18. Improving the Stability and Efficiency of Diffusion Models for Content Consistent Super-Resolution / L. Sun et al. IEEE Transactions on Image Processing. 2025. Vol. 34. P. 8421–8434. URL: https://doi.org/10.1109/tip.2025.3640863 (Last accessed: 22.02.2026).

19. Yue Z., Wang J., Loy C. C. Efficient Diffusion Model for Image Restoration by Residual Shifting. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2025. Vol. 47, no. 1. P. 116–130. URL: https://doi.org/10.1109/tpami.2024.3461721 (Last accessed: 22.02.2026).

20. SinSR: Diffusion-Based Image Super-Resolution in a Single Step / Y. Wang et al. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 16–22 June 2024. 2024. P. 25796–25805. URL: https://doi.org/10.1109/cvpr52733.2024.02437 (Last accessed: 22.02.2026).

21. Salimans T., Ho J. Progressive Distillation for Fast Sampling of Diffusion Models. *International Conference on Learning Representations*. 2022. URL: https://openreview.net/forum?id=TIdIXIpzhoI (Last accessed: 22.02.2026).

22. One-Step Effective Diffusion Network for Real-World Image Super-Resolution / Wu R., Sun L., Ma Z., Zhang L. *Advances in Neural Information Processing Systems*. 2024. Vol. 37. P. 92529–92553. URL: https://openreview.net/forum?id=TPtXnpRvur (Last accessed: 22.02.2026).

23. ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation / Z. Wang et al. *Proceedings of the 37th International Conference on Neural Information Processing Systems*, New Orleans, LA, USA. 2023. URL: https://dl.acm.org/doi/10.5555/3666122.3666490 (Last accessed: 22.02.2026).

24. TSD-SR: One-Step Diffusion with Target Score Distillation for Real-World Image Super-Resolution / L. Dong et al. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 10–17 June 2025. 2025. P. 23174–23184. URL: https://doi.org/10.1109/cvpr52734.2025.02158 (Last accessed: 22.02.2026).

25. AddSR: Accelerating Diffusion-based Blind Super-Resolution with Adversarial Diffusion Distillation / R. Xie et al. Pattern Recognition. 2026. Vol. 175. P. 113012. URL: https://www.sciencedirect.com/science/article/pii/S0031320325016759 (Last accessed: 22.02.2026).

26. Recognize Anything: A Strong Image Tagging Model / Y. Zhang et al. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 17–18 June 2024. 2024. P. 1724–1732. URL: https://doi.org/10.1109/cvprw63382.2024.00179 (Last accessed: 22.02.2026).

27. Learning Transferable Visual Models From Natural Language Supervision / A. Radford et al. *Proceedings of the 38th International Conference on Machine Learning. 2021*. Vol. 139. P. 8748–8763. URL: https://proceedings.mlr.press/v139/radford21a.html (Last accessed: 22.02.2026).

28. Taming Diffusion Prior for Image Super-Resolution with Domain Shift SDEs / Q. Cui et al. *Proceedings of the 38th International Conference on Neural Information Processing Systems*, Vancouver, BC, Canada. 2024. URL: https://dl.acm.org/doi/10.5555/3737916.3739271 (Last accessed: 22.02.2026).

29. Yue Z., Liao K., Loy C. C. Arbitrary-steps Image Super-resolution via Diffusion Inversion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2025. P. 23153–23163. URL: https://openaccess.thecvf.com/content/CVPR2025/html/Yue_Arbitrary-steps_Image_Super-resolution_via_Diffusion_Inversion_CVPR_2025_paper.html (Last accessed: 22.02.2026).

30. Agustsson E., Timofte R. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, 21–26 July 2017. 2017. URL: https://doi.org/10.1109/cvprw.2017.150 (Last accessed: 22.02.2026).

31. Toward Real-World Single Image Super-Resolution: A New Benchmark and a New Model / J. Cai et al. *2019 IEEE/CVF International Conference on Computer Vision (ICCV),* Seoul, Korea (South), 27 October – 2 November 2019. 2019. URL: https://doi.org/10.1109/iccv.2019.00318 (Last accessed: 22.02.2026).

32. Component Divide-and-Conquer for Real-World Image Super-Resolution / P. Wei et al. *Computer Vision – ECCV 2020*. Cham, 2020. P. 101–117. URL: https://doi.org/10.1007/978-3-030-58598-3_7 (Last accessed: 22.02.2026).

33. Image Quality Assessment: From Error Visibility to Structural Similarity / Z. Wang et al. IEEE Transactions on Image Processing. 2004. Vol. 13, no. 4. P. 600–612. URL: https://doi.org/10.1109/tip.2003.819861 (Last accessed: 22.02.2026).

34. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric / R. Zhang et al. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 18–23 June 2018. 2018. URL: https://doi.org/10.1109/cvpr.2018.00068 (Last accessed: 22.02.2026).

35. Wang J., Chan K. C. K., Loy C. C. Exploring CLIP for Assessing the Look and Feel of Images. *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*. 2023. URL: https://dl.acm.org/doi/10.1609/aaai.v37i2.25353 (Last accessed: 22.02.2026).

36. MUSIQ: Multi-scale Image Quality Transformer / J. Ke et al. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 10–17 October 2021. 2021. URL: https://doi.org/10.1109/iccv48922.2021.00510 (Last accessed: 22.02.2026).

37. Adversarial Diffusion Compression for Real-World Image Super-Resolution / B. Chen et al. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 10–17 June 2025. 2025. P. 28208–28220. URL: https://doi.org/10.1109/cvpr52734.2025.02627 (Last accessed: 22.02.2026).

38. Image Quality Assessment: Unifying Structure and Texture Similarity / Ding K., Ma K., Wang S., Simoncelli E. P. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022. Vol. 44, no. 5. P. 2567–2581. URL: https://doi.org/10.1109/tpami.2020.3045810 (Last accessed: 22.02.2026).

39. PassionSR: Post-Training Quantization with Adaptive Scale in One-Step Diffusion based Image Super-Resolution / L. Zhu et al. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 10–17 June 2025. 2025. P. 12778–12788. URL: https://doi.org/10.1109/cvpr52734.2025.01192 (Last accessed: 22.02.2026).

40. Edge-SD-SR: Low Latency and Parameter Efficient On-device Super-Resolution with Stable Diffusion via Bidirectional Conditioning / M. Noroozi et al. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 10–17 June 2025. 2025. P. 12789–12798. URL: https://doi.org/10.1109/cvpr52734.2025.01193 (Last accessed: 22.02.2026).

41. BiMaCoSR: Binary One-Step Diffusion Model Leveraging Flexible Matrix Compression for Real Super-Resolution / K. Liu et al. *Proceedings of the 42nd International Conference on Machine Learning,* Vancouver, Canada. 2025. URL: https://dl.acm.org/doi/abs/10.5555/3780338.3781942 (Last accessed: 22.02.2026).

Стаття надійшла до редакції 25.03.2026.

Стаття пройшла рецензування 29.03.2026.

Стаття опублікована 31.03.2026.

Козлов Сергій Леонідович – аспірант кафедри комп'ютерних наук, ORCID: 0009-0000-5772-1085, e-mail: serhii.kozlov@gmail.com.

Колесницький Олег Костянтинович – канд. техн.наук, професор кафедри комп'ютерних наук, ORCID: 0000-0003-0336-4910.

Вінницький національний технічний університет.