

**Теглівець О.В.**

Національний університет «Львівська політехніка»

## ВПЛИВ СПОТВОРЕНОЇ КОНТРАСТНОСТІ НА СПРИЙНЯТТЯ ЯКОСТІ ЗОБРАЖЕНЬ: ПОРІВНЯННЯ ОЦІНОК ЕКСПЕРТІВ ТА БАЗИ TID2013

У статті розглянуто проблему суб'єктивного сприйняття якості цифрових зображень зі спотвореною контрастністю. У цифрових зображеннях контрастність є одним із найбільш чутливих і варіативних типів візуальних артефактів. Метою статті є порівняння суб'єктивних оцінок якості зображень зі спотвореною контрастністю, отриманих від експертної групи в реальних умовах перегляду, із відповідними значеннями MOS бази TID2013 для встановлення рівня кореляції, варіативності та статистичної узгодженості між двома системами оцінювання.

Проведено експериментальне порівняння середніх суб'єктивних оцінок (MOS), отриманих від групи з 30 експертів у реальних (неконтрольованих) умовах перегляду (різні монітори, ноутбуки, смартфони, різне освітлення та відстань до екрана), з еталонними значеннями MOS бази TID2013 – однієї з найавторитетніших і найбільш цитованих баз даних для оцінки якості зображень. Для дослідження обрано п'ять зображень із набору TID2013 (i01\_17\_1 – i01\_17\_5), що мають різні рівні спотворення контрастності (від помірного до критичного). Кожен експерт оцінював якість за дев'ятибальною шкалою (0 – найгірша якість, 9 – найкраща). На основі індивідуальних оцінок розраховано MOS для кожного зображення (діапазон 3.233–6.833).

Проведений парний *t*-тест та аналіз різниць виявив статистично значуще систематичне зміщення: експерти в реальних умовах в середньому ставили на 0.5500544 бала вище, ніж вибірка TID2013. При цьому ранговий порядок п'яти зображень за рівнем сприйнятої якості в обох наборах повністю збігся, що підтверджує високу внутрішню узгодженість і надійність бази TID2013 навіть для контрастних спотворень.

Виявлене позитивне зміщення можна пояснити адаптацією зорової системи до типових умов повсякденного перегляду, коли користувачі мени критично ставляться до помірних відхилень контрастності порівняно з контрольованим лабораторним середовищем. Це має важливі практичні наслідки для сучасних завдань глибокого навчання: моделі покращення та відновлення зображень (MIRNet, HDR-Net, Zero-DCE, RetinexNet тощо), навчені на MOS TID2013 без корекції, систематично недооцінюватимуть якість результату з погляду реального користувача. Запропоновано низку практичних рекомендацій: додавати сталу корекцію 0.5500544 бала до всіх цільових MOS TID2013 при навчанні моделей; інтегрувати оцінки, зібрані в неконтрольованих умовах, як додатковий компонент функції втрат, застосовувати перцептивні ваги, розширювати тренувальні набори комбінацією лабораторних і реальних суб'єктивних оцінок.

**Ключові слова:** контрастність, MOS, база TID2013, глибоке навчання, оцінка якості зображень.

**Постановка проблеми.** Якість цифрових зображень на сьогодні є надзвичайно важливою для цифрової фотографії, мультимедійних сервісів, медицини, відеомоніторингу тощо. При цьому треба не лише зберегти формальні характеристики сигналу, а й забезпечити відповідність якості зображення для людського сприйняття. Однією з основних характеристик, яка безпосередньо пов'язана із суб'єктивною оцінкою якості зображення, є контрастність. Проблема кількісного оцінювання якості цифрових зображень у контексті спотворення контрастності посідає важливе місце в сучасних дослідженнях

комп'ютерного зору та цифрового оброблення сигналів, оскільки контрастність є багатовимірним поняттям. На практиці вона залежить від багатьох факторів: умов освітлення, характеристик сенсора, обраної колірної моделі, психофізичних особливостей експерта тощо.

Оскільки людина часто є кінцевим споживачем візуальних продуктів (відео, зображення), суб'єктивна оцінка часто найбільше підходить, щоб отримати кількісне значення якості зображення за певною шкалою значень. Однак таку оцінку не можна безпосередньо використати

в автоматизованому обробленні зображень, оскільки вона залежить від особливостей зорової системи людини [1].

Сучасні системи оцінювання якості цифрових зображень базуються як на об'єктивних метриках (PSNR, SSIM, FSIM, LPIPS), так і на суб'єктивних методах, серед яких найпоширенішим є MOS (Mean Opinion Score), і перцептивні показники, такі як LPIPS [2-3]. Саме суб'єктивне оцінювання відображає реакцію сприйняття людського візуального апарату, проте воно може суттєво варіювати залежно від складу групи експертів, умов тестування, попереднього досвіду та індивідуальних особливостей [4]. Тому актуальною є перевірка надійності й узгодженості вже наявних баз даних шляхом їхньої незалежної валідації.

Із розвитком математичних, алгоритмічних та апаратних засобів зростає сфера застосування глибоких нейронних мереж для оцінки та покращення контрастності зображень. Візуальні експертні оцінки можуть бути використані як цільові функції в процесі навчання нейронних мереж. Це зумовлює необхідність формування та використання спеціальних баз даних, у яких кожен елемент (зображення, відео, мовний сигнал) має не лише еталонну версію та різні типи спотворень, але й зареєстровані MOS великої групи експертів. Такі бази виконують подвійну функцію: з одного боку, вони забезпечують дослідників стандартним набором даних для порівняння алгоритмів, а з іншого – слугують тренувальним матеріалом для побудови моделей об'єктивної оцінки якості, що апроксимують MOS (Mean Opinion Score) [4-6].

Однією з найбільш використовуваних у наукових дослідженнях є база TID2013, яка містить 3000 зображень із 24 типами спотворень та суб'єктивними MOS-оцінками, зібраними за участі понад 900 анонімних оцінювачів із різних країн [6]. Незважаючи на масштабність та репрезентативність вибірки, питання надійності оцінок для окремих категорій спотворень, зокрема контрастності, залишається відкритим. Тобто контрастність є особливим типом спотворення, оскільки її не завжди рівномірно сприймають різні групи експертів, що часто демонструють різну чутливість до змін локальної чи глобальної контрастності [7]. Надмірне зниження або, навпаки, підвищення контрастності веде до втрати деталей, спотворення відносних відмінностей між об'єктами й зрештою до зниження сприйманої якості. Саме тому контрастні спотворення активно моделюються в стандартних базах тестових зображень, зокрема в базі TID2013 [5-6].

У цьому контексті важливим є питання, чи будуть MOS, отримані від невеликої групи експертів, узгоджуватися з MOS, наданими широкою аудиторією спеціалістів, як у TID2013. Це питання набуває особливої актуальності для завдань, де помилки у сприйнятті контрастності можуть мати практичні наслідки (наприклад, у медичній візуалізації чи контролі промислових процесів).

Порівняння таких оцінок дає змогу встановити, наскільки MOS TID2013 дійсно відображає сприйняття якості зображень із контрастним спотворенням та наскільки цю базу можна буде використовувати як основу в подальших дослідженнях. При цьому важливо не лише порівняти середні значення, але й дослідити статистичні характеристики варіативності, які відображають стабільність і надійність оцінювання [8].

Сучасний етап розвитку досліджень характеризується широким використанням методів глибокого навчання для побудови моделей об'єктивної оцінки якості зображень, які безпосередньо навчаються прогнозувати MOS за даними зображення [9-10]. MOS використовують як цільову функцію у завданнях оптимізації, де, наприклад, параметри кодека або алгоритму покращення якості підбирають таким чином, щоб максимізувати очікуваний MOS експертів. Також MOS відіграє роль певного стандарту, на якому навчають нейронну мережу, оскільки без надійних баз із MOS неможливо побудувати і валідувати адекватні об'єктивні моделі якості.

У практиці наукових досліджень та інженерних застосувань MOS зазвичай супроводжується додатковим аналізом, спрямованим на перевірку статистичної значущості відмінностей між різними умовами тестування. Зокрема поширеним є використання t-тесту для парних або незалежних вибірок для виявлення груп умов, що істотно відрізняються між собою за MOS [11]. Такі підходи дозволяють уникнути хибних висновків, коли незначні відмінності в середніх значеннях MOS інтерпретують як суттєві, хоча насправді вони можуть бути зумовлені випадковими флуктуаціями в оцінках спостерігачів.

**Аналіз останніх досліджень і публікацій.** Питання взаємозв'язку між спотворенням контрастності та перцептивною якістю порушували в багатьох класичних і сучасних дослідженнях. У [12], одному з перших фундаментальних досліджень, розкрито принципи сприйняття контрастності та обґрунтовано, що зорове сприйняття залежить не лише від абсолютних значень яскравості, а й від локальних співвідношень інтенсивностей.

Ця концепція лягла в основу сучасних моделей локальної адаптації, які пояснюють, чому навіть незначні відхилення контрастності можуть суттєво вплинути на візуальне враження від зображення.

У [13-14] визначено низку критичних факторів, що впливають на сприйняття якості зображення: просторові частоти, глибина тіней, характер переходів між світлими і темними ділянками. У [15-16] показано, що feature-based metrics (SSIM, FSIM, GMSD) демонструють кращу відповідність суб'єктивним оцінкам, але все ще не забезпечують абсолютно точної кореляції під час зміни контрастності. У [17] зауважено, що навіть невеликі зміни контрастності можуть суттєво змінювати оцінку загальної якості та інформативності зображення.

У [18] зазначено, що моделі, орієнтовані на перцептивну якість, вимагають достовірних даних, що є максимально наближеними до реального людського сприйняття. Тому якість MOS у тренувальних і валідаційних наборах сучасних моделей HDR-Net, MIRNet, Zero-DCE, RetinexNet має суттєве значення.

Порівняння MOS, отриманих від експертів у реальних умовах перегляду, з MOS, представленими в TID2013, розглянуто лише в поодиноких дослідженнях. Зокрема, у [19] звернено увагу на проблему індивідуальних профілів сприйняття, що дозволяє пояснити розбіжності між оцінками різних груп оцінювачів. У [20] наголошено на тому, що неконтрольовані умови перегляду та індивідуальні особливості зору оцінювачів значно впливають на MOS. Це свідчить про важливість аналізу кореляції та варіативності між двома системами оцінювання.

**Постановка завдання.** Метою статті є порівняння суб'єктивних оцінок якості зображень зі спотвореною контрастністю, отриманих від експертної групи в реальних умовах перегляду, із відповідними значеннями MOS бази TID2013 для встановлення рівня кореляції, варіативності та статистичної узгодженості між двома системами оцінювання.

**Виклад основного матеріалу.** Особливість нашого дослідження полягала в тому, що умови,

в яких група експертів оцінювала якість зображення, були наближеними до реального використання, зокрема це стосувалося й екранів пристроїв із різною роздільною здатністю (смартфони, різні монітори комп'ютерів чи екрани ноутбуків). У [8] наголошено на тому, що якісні експерименти рідко проводять у контрольованих умовах. Для оцінювання обрано зображення із бази TID2013 з різним рівнем спотворення контрастності: зображення i01\_17\_1, i01\_17\_2, i01\_17\_3, i01\_17\_4, i01\_17\_5 (рис. 1).

Для оцінювання було залучено групу з 30 експертів, із них 11 жінок і 19 чоловіків. Вибірка з 30 експертів забезпечує достатню кількість оцінок для MOS і кореляції з PSNR/SSIM, а також базується на статистичній надійності та практиках TID2013.

Експертні оцінки нормалізували в межах кожного базового зображення (5 варіантів зображення). Кожен експерт оцінював якість тестових зображень за дев'ятибальною шкалою, де 0 відповідав найгіршій якості зображення, а 9 – найкращій. На основі отриманих оцінок було розраховано середнє значення MOS, що відображає узагальнену суб'єктивну оцінку якості зображення та дає змогу порівнювати результати для різних типів спотворень. MOS визначають за формулою:

$$MOS = \frac{1}{N} \cdot \sum_{i=1}^N R_i, \text{ де} \quad (1)$$

$N$  – загальна кількість експертів або оцінок (розмір вибірки);

$R_i$  – індивідуальна оцінка, яку надав окремий експерт;

$\sum_{i=1}^N R_i$  – сума всіх оцінок, які надали експерти;

$\frac{1}{N}$  – коефіцієнт нормалізації, який ділить суму на кількість оцінку.

Для зображення i01\_17\_1 експерти дали такі оцінки для загальної якості зображення: 6, 7, 7, 5, 7, 9, 4, 8, 6, 4, 9, 7, 7, 6, 6, 7, 5, 9, 8, 4, 7, 7, 6, 5, 5, 6, 9, 3, 8, 6, відповідно:

$$MOS_{\text{експ.}} = \frac{(3 \times 1 + 4 \times 3 + 5 \times 4 + 6 \times 7 + 7 \times 8 + 8 \times 3 + 9 \times 4)}{30} = \frac{193}{30} = 6.433$$

MOS=6.433 вказує на те, що більшість експертів оцінила зображення як прийнятне, тобто таке, що не має критичних спотворень, які б могли вплинути



Рис. 1. Приклади зображень зі спотвореною контрастністю із бази TID2013

на загальне сприйняття зображення. Більшість оцінок (53 %) зосереджено в зоні 6-7. Є потенціал для стабілізації MOS  $\geq 7.0$  при STD  $\leq 1.3$ .

Для зображення i01\_17\_2 експерти дали такі оцінки: 7, 7, 4, 9, 7, 8, 5, 9, 8, 6, 8, 8, 9, 5, 7, 2, 5, 8, 6, 8, 8, 7, 8, 6, 7, 6, 6, 7, 8, 6, відповідно:

$$MOS_{\text{експ.}} = \frac{(2 \times 1 + 4 \times 1 + 5 \times 3 + 6 \times 6 + 7 \times 7 + 8 \times 9 + 9 \times 3)}{30} = \frac{205}{30} = 6.833$$

MOS=6.833 вказує на те, що експерти сприймають зображення як прийнятне з вираженою тенденцією до високої якості. Більшість оцінок (67 %) зосереджено в зоні 6-8, з піком на 7-8. Низькі оцінки (2-5) становлять лише 13 %, високі (9) – 7 %. Є потенціал для стабілізації MOS  $\geq 7.0$ -7.5 при SEM  $\leq 0.25$ .

Для зображення i01\_17\_3 експерти дали такі оцінки для загальної якості зображення: 4, 5, 5, 8, 8, 5, 4, 7, 4, 3, 6, 5, 4, 4, 5, 8, 3, 6, 5, 3, 6, 6, 7, 3, 4, 7, 7, 2, 7, 4, відповідно:

$$MOS_{\text{експ.}} = \frac{(2 \times 1 + 3 \times 4 + 4 \times 7 + 5 \times 6 + 6 \times 4 + 7 \times 5 + 8 \times 3)}{30} = \frac{155}{30} = 5.167$$

MOS=5.167 вказує на те, що у вибірці домінує середній рівень сприйняття якості зображення. Низькі оцінки ( $\leq 4$ ) становлять 30 %, високі ( $\geq 7$ ) – лише 23 %. Це означає, що навіть при повторних тестах MOS з високою ймовірністю не перевищить 6.0. Є потенціал для стабілізації MOS до  $\geq 6.5$  при STD  $\leq 1.3$  та відсутності оцінок  $\leq 4$ .

Для зображення i01\_17\_4 експерти дали такі оцінки для загальної якості зображення: 7, 6, 3, 9, 8, 4, 7, 7, 9, 9, 6, 9, 9, 3, 8, 4, 6, 4, 7, 2, 7, 7, 6, 6, 8, 7, 5, 9, 7, 7, відповідно:

$$MOS_{\text{експ.}} = \frac{(2 \times 1 + 3 \times 2 + 4 \times 3 + 5 \times 1 + 6 \times 5 + 7 \times 9 + 8 \times 3 + 9 \times 6)}{30} = \frac{196}{30} = 6.533$$

MOS=6.533 вказує на те, що експерти загалом сприймають якість зображення як прийнятну, але з помітною варіативністю оцінок. Більшість оцінок (53 %) зосереджено в зоні 6-7, з піком на 7, але присутні значні відхилення: низькі оцінки ( $\leq 4$ ) становлять 20 %, високі (9) – 17 %. Є потенціал для стабілізації MOS  $\geq 7.0$  при STD  $\leq 1.3$  та відсутності оцінок  $\leq 4$ .

Для зображення i01\_17\_5 експерти дали такі оцінки для загальної якості зображення: 2, 4, 1, 2, 4, 1, 2, 5, 3, 4, 5, 4, 2, 2, 4, 4, 3, 4, 5, 1, 2, 5, 5, 1, 2, 7, 1, 1, 8, 3, відповідно:

$$MOS_{\text{експ.}} = \frac{(1 \times 6 + 2 \times 7 + 3 \times 3 + 4 \times 7 + 5 \times 5 + 7 \times 1 + 8 \times 1)}{30} = \frac{97}{30} = 3.233$$

MOS=3.233 вказує на те, що експерти сприймають якість зображення як дуже низьку. Високі оцінки ( $\geq 7$ ) становлять лише 6.7 %, а оцінки 5 – 16.7 %. Є потенціал для стабілізації MOS  $\geq 5.5$  при STD  $\leq 1.5$  та відсутності оцінок  $\leq 3$ .

Для порівняння оцінок застосовано парний t-тест – статистичний метод, який використовують

для порівняння середніх значення двох пов'язаних груп. Такий підхід дозволяє уникнути хибних висновків, коли незначні відмінності в середніх значеннях MOS інтерпретують як суттєві, хоча насправді вони можуть бути зумовлені випадковими флуктуаціями в оцінках спостерігачів. Відповідно для кожної пари оцінок було обчислено різницю ( $d$ ) (Таблиця 1). Парний t-тест використовує різниці  $d$ , щоб перевірити, чи середня різниця  $\bar{d}$  є статистично значущою, тобто чи не виникла вона випадково.

Таблиця 1

Порівняльний аналіз результатів MOS TID2013 та MOS експертів

Файл	MOS TID2013	MOS експерти	Різниця ( $d$ )
i01_17_1	5,53135	6,433	0.90165
i01_17_2	6,62162	6,833	0.21138
i01_17_3	4,08108	5,167	1.085912
i01_17_4	6,24324	6,533	0.28976
i01_17_5	2,97143	3,233	0.26157

Середнє значення різниці становить:

$$\bar{d} = \frac{0.90165 + 0.21138 + 1.085912 + 0.28976 + 0.26157}{5} = \frac{2.750272}{5} = 0.5500544$$

Як бачимо з дослідження, суб'єктивне сприйняття якості цифрових зображень зі спотвореною контрастністю є надзвичайно чутливим типом візуального спотворення. Порівняльний аналіз MOS, отриманих від 30 кваліфікованих експертів у реальних (неконтрольованих) умовах перегляду, з еталонними значеннями MOS бази TID2013 виявив систематичне зміщення: експерти стабільно оцінюють якість приблизно на 0.5500544 бала вище, ніж учасники лабораторного тестування TID2013 (рис. 2).



Рис. 2. Порівняння показників MOS TID2013 і MOS експертів

Цей результат має особливе значення для розуміння відмінностей між лабораторними та реальними умовами сприйняття. Слід зауважити, що ранговий порядок п'яти тестових зображень за

рівнем сприйнятої якості практично збігся з порядком у базі TID2013. Для сучасних архітектур щодо покращення зображень, зокрема таких як MIRNet, використання MOS із TID2013 як цільових значень у функції втрат призводитиме до недооцінки якості на виході моделі з погляду реального експерта. Тому перед тренування можна рекомендувати додавання +0.5500544 бала до всіх цільових MOS TID2013. Також до тренувального набору можна включити нові експертні MOS (зібрані в реальних умовах), ввести вагові коефіцієнти для різних типів контрастних спотворень на основі виявленої чутливості експертів, застосовувати подвійне валідування, поєднуючи лабораторні MOS та реальні експертні оцінки, розширити саму вибірку зображень для кожного типу спотворень.

**Висновки.** Проведене дослідження показало, що суб'єктивне сприйняття якості зображень зі спотвореною контрастністю є надзвичайно чутливим і варіативним типом спотворення. Порівняння MOS, отриманих від групи з 30 експертів, із відповідними значеннями бази TID2013 виявило статистично значущу систематичну різницю: експерти в реальних (неконтрольованих) умовах перегляду в середньому оцінюють такі зображення на 0.5500544 бала вище.

Водночас напрямок зміни якості зображень в обох системах оцінювання виявився практично ідентичним. Це підтверджує високу надійність бази TID2013 як еталонного набору даних навіть

для такого складного типу спотворень, як контрастність.

Основні статистичні характеристики експертних оцінок (MOS=3.233–6.833) свідчать про помірну внутрішню варіативність і достатню статистичну надійність вибірки  $n=30$  для виявлення значних відмінностей.

Отримані результати мають важливе практичне значення: база TID2013 залишається репрезентативним джерелом суб'єктивних оцінок для контрастних спотворень, але під час використання в завданнях навчання або валідації моделей покращення якості зображень доцільно враховувати систематичне зміщення +0.5500544 бала, щоб наблизити до оцінок експертів у реальних умовах перегляду. Неконтрольовані умови перегляду (різні дисплеї, освітлення, відстань) не руйнують загальної узгодженості оцінок, а лише зсувають MOS, що робить експерименти, проведені в реальних умовах, цінним джерелом даних для адаптації моделей до реального користувацького досвіду. Для подальшого підвищення перцептивної якості моделей типу MIRNet рекомендовано: інтегрувати експертні оцінки, отримані в неконтрольованих умовах, як додатковий компонент функції втрат; застосовувати зміщення MOS або перцептивні ваги, що враховують виявлене систематичне перевищення, розширювати тренувальні набори комбінацією лабораторних і реальних суб'єктивних оцінок.

#### Список літератури:

1. Talun A., Drozda P., Yelmanov S., Romanyshyn Y., Tehlivets O. Convolutional Neural Network Assessment of Image Quality Based on the TID2013 Database. *IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS*. 2023. Pp. 1088–1092. <https://doi.org/10.1109/IDAACS58523.2023.10348876>
2. Wang Z., Bovik A.C., Sheikh H. R., Simoncelli E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004. 13 (4). Pp. 600–612. <https://doi.org/10.1109/TIP.2003.819861>
3. Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. Pp. 586–595.
4. Sheikh, H.R., Sabir, M.F. and Bovik, A.C. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*. 2006. 15. Pp. 3440–3451. <https://doi.org/10.1109/TIP.2006.881959>
5. Ponomarenko N., Ieremeiev O., Lukin V., Jin L., Egiazarian K., Astola J., Vozel B., Chehdi K., Carli M., Battisti F., Kuo C.-C. J. A new color image database TID2013: Innovations and results. *Proceedings of ACIVS, Poznan, Poland, Oct.* 2013. Pp. 402–413.
6. Ponomarenko N., Jin L., Ieremeiev O., Lukin V., Egiazarian K., Astola J., Vozel B., Chehdi K., Carli M., Battisti F., Kuo, C.-C. J. Image database TID2013: Peculiarities, results and perspectives, *Signal Processing: Image Communication*, 2015. 30. Pp. 57–77 <https://doi.org/10.1016/j.image.2014.10.009>
7. Wang Z., Bovik A.C. Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, 26. 2009. Pp. 98–117. <https://doi.org/10.1109/MSP.2008.930649>
8. Mantiuk R.K.; Tomaszewska A.; Mantiuk R. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum*, 31(8). 2012. Pp. 2478–2491. <https://doi.org/10.1111/j.1467-8659.2012.03188.x>

9. Bosse S., Maniry D., Muller K.-R., Wiegand T., Samek, W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing*. 2018. 27(1), 206–219.
10. Talebi H., Milanfar P. NIMA: Neural image assessment. *IEEE Transactions on Image Processing*. 2018. 27(8). Pp. 3998–4011. <https://doi.org/10.1109/TIP.2018.2831899>
11. Hossfeld T., Tran-Gia Ph., Fiedler M. Quantification of Quality of Experience for Edge-Based Applications. 2007. 4516. 361–373. [https://doi.org/10.1007/978-3-540-72990-7\\_34](https://doi.org/10.1007/978-3-540-72990-7_34)
12. Land E.H., McCann J.J. Lightness and retinex theory. *J Opt Soc Am*. 1971 Jan; 61(1). 1–11. <https://doi.org/10.1364/josa.61.000001>
13. Watson A.B. Efficiency of a model human image code. *J Opt Soc Am A*. 1987 Dec. 4(12). 2401–17. <https://doi.org/10.1364/josaa.4.002401>
14. Keelan B. *Handbook of Image Quality*. Marcel Dekker. 2002
15. Zhang L. Zhang X. Mou and D. Zhang. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing*. 2011. 20(8). Pp. 2378–2386. <https://doi.org/10.1109/TIP.2011.2109730>
16. Xue W., Zhang L., Mou X. Bovik A. C. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index. *IEEE Transactions on Image Processing*. 2014. 23(2). Pp. 684–695. <https://doi.org/10.1109/TIP.2013.2293423>
17. Peli E. Contrast in complex images. *Journal of the Optical Society of America A, Optics and image science*, 2020. 7(10). Pp. 2032–2040. <https://doi.org/10.1364/josaa.7.002032>. PMID: 2231113
18. Zamir S.W., Arora A., Khan S., Hayat M., Khan F.S., Yang M.-H. Shao L. Learning Enriched Features for Real Image Restoration and Enhancement. *Vedaldi, A., Bischof, H., et al., Eds., European Conference on Computer Vision, Springer, Cham*. 2020. 12370. 492–511. [https://doi.org/10.1007/978-3-030-58595-2\\_30](https://doi.org/10.1007/978-3-030-58595-2_30)
19. Ding K., Ma K., Wang S., Simoncelli E.P. Image Quality Assessment: Unifying Structure and Texture Similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020. 44. 2567–2581
20. Winkler S. *Digital video quality: Vision models and metrics*. 2013. 200 p.

#### **Tehlivets O.V. THE IMPACT OF DISTORTED CONTRAST ON THE PERCEIVED QUALITY OF IMAGES: COMPARISON OF EXPERT ASSESSMENTS WITH THE TID2013 DATABASE**

*The article examines the problem of subjective perception of digital image quality in the presence of contrast distortion. In digital imaging, contrast is one of the most sensitive and variable types of visual artifacts. The aim of the article is to compare subjective quality assessments of contrast-distorted images provided by an expert group under real viewing conditions with the corresponding MOS values from the TID2013 database in order to determine the level of correlation, variability, and statistical consistency between the two evaluation systems.*

*An experimental comparison was conducted between the mean subjective scores (MOS) obtained from a group of 30 experts in real (uncontrolled) viewing conditions (different monitors, laptops, smartphones, various lighting conditions and viewing distances) and the reference MOS values from TID2013 – one of the most authoritative and widely cited image quality assessment databases. Five images from TID2013 (i01\_17\_1 – i01\_17\_5) with different levels of contrast distortion (from moderate to severe) were selected for the study. Each expert evaluated image quality using a nine-point scale (0 = worst quality, 9 = best quality). Based on the individual ratings, MOS values for each image (range 3.233–6.833).*

*The paired t-test and difference analysis revealed a statistically significant systematic shift: on average, experts under real-world conditions rated the images 0.5500544 points higher than the TID2013 scores. At the same time, the ranking order of the five images by perceived quality was identical in both datasets, confirming the strong internal consistency and reliability of TID2013 even for contrast distortions.*

*The observed positive shift can be explained by the adaptation of the human visual system to typical everyday viewing conditions, where users tend to be less critical of moderate contrast deviations compared to controlled laboratory environments. This has important practical implications for modern deep learning tasks: enhancement and restoration models (MIRNet, HDR-Net, Zero-DCE, RetinexNet, etc.) trained solely on TID2013 MOS without correction will systematically underestimate perceived quality from the real user's perspective. Several practical recommendations are proposed: adding a constant correction of 0.5500544 points to all TID2013 target MOS during model training; integrating ratings collected under uncontrolled conditions as an auxiliary component of the loss function; applying perceptual weighting; and expanding training datasets through a combination of laboratory and real subjective evaluations.*

**Key words:** contrast, MOS, TID2013 database, deep learning, image quality assessment.

Дата надходження статті: 21.11.2025

Дата прийняття статті: 10.12.2025

Опубліковано: 30.12.2025