



Univerzitet u Nišu
Elektronski fakultet
Katedra za računarstvo



Super-resolution

Tehnički izveštaj

Mentor:

prof. dr Aleksandar Milosavljević

Studenti:

Petar Stokić 1408
Darjan Drugarinović 1303

Niš, 2022

Sadržaj

1. Uvod	1
2. Ocena kvaliteta videa	2
3. Metode video super rezolucije	3
3.1 Metode sa poravnanjem	4
3.1.1 Metode procene pokreta i kompenzacije	4
3.2 Metode bez poravnanja	5
3.2.1 2D konvolucione metode	5
3.2.2 3D konvolucione metode	5
3.2.3 Rekurentne konvolucione neuronske mreže (RCNNs)	5
3.2.4 Ne-lokalne metode	5
3.2.4 Druge metode	5
4. Realizacija sistema namenjenog super rezoluciji videa	6
5. SRGAN	7
5.1 Povezani radovi	7
5.1.1 Super rezolucija slike	7
5.1.2 Dizajn konvolucionih neuronskih mreža	8
5.2 Adversarialna mreža	9
5.3 Perceptualna funkcija gubitaka	10
5.4 Poređenje sa drugim modelima	10
6. ESRGAN	12
6.1 Arhitektura mreže	13
6.2 Relativistički diskriminator	14
6.3 Perceptualni gubitak	14
6.4 Mrežna interpolacija	15
6.5 Rezultati ESRGAN modela	15
7. Real-ESRGAN	20
7.1 Degradacija slika	20
7.2 Model degradacije visokog reda	21
7.3 Artefakti zvonjave i prekoračenja	22
7.4 Mreže i obuka	23
7.4.1 ESRGAN generator	23
7.4.2 U-Net diskriminator sa spektralnom normalizacijom (SN)	24
7.4.3 Proces obuke	24

7.5 Rezultati Real-ESRGAN modela	25
7.6 Ograničenja	26
8. Zaključak	27

1. Uvod

Cilj super-rezolucije (SR, eng. Super Resolution) je da generiše sliku ili video zapis visoke rezolucije (HR, eng. High Resolution) iz odgovarajuće slike ili video zapisa niske rezolucije (LR, eng. Low Resolution). [1] Ukoliko se materijal generiše na osnovu slike niske rezolucije, reč je o “Image Super Resolution” (ISR). U slučaju video zapisa, ovaj proces se naziva “Video Super Resolution” (VSR).

Metode koje se primenjuju kod ISR vrše obradu jedne slike, dok VSR metode koriste više uzastopnih slika, odnosno okvira ili frejmova (eng. Frame), da bi se obavila super rezolucija jednog oređenog frejma.

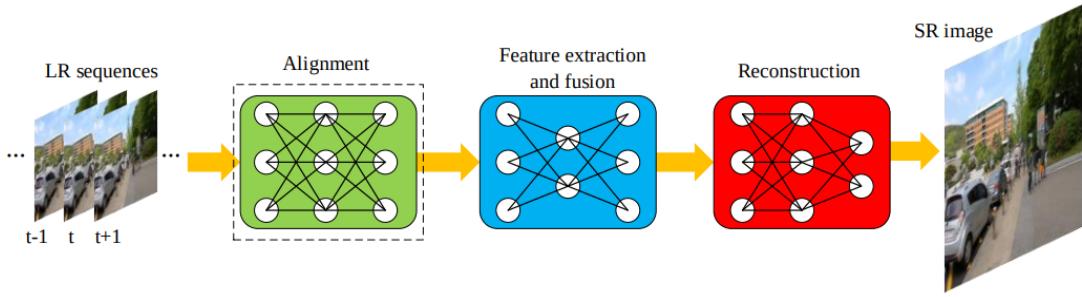
Video je jedan od najčešćih multimedijalnih formata, pa je samim tim i njegova super-rezolucija veoma važna. Sa napretkom u tehnologiji prikaza, VSR postaje sve značajniji za video snimke niske rezolucije. Na primer, sve veći broj ekrana može da emituje video snimke ultra visoke rezolucije, 4K (3840×2160) i 8K (7680×4320), ali sadržaji koji odgovaraju toj rezoluciji su još uvek retki. Super rezolucija je našla primenu i u mnogim drugim oblastima, kao što su medicinska ili satelitska snimanja.

U širem smislu, VSR se može smatrati proširenjem ISR, i može se postići primenom ISR algoritama nad pojedinačnim frejmovima, umesto na više uzastopnih frejmova. Međutim, ISR tehnike mogu uneti artefakte, što utiče na vremensku nekoherentnost u video zapisu. Kako performanse ISR nisu uvek zadovoljavajuće, za VSR se koriste tehnike koje se oslanjaju na veći broj frejmova.

Poslednjih godina predloženo je mnogo algoritama video super rezolucije, koji se dele u dve kategorije: tradicionalne metode i metode zasnovane na dubokom učenju. Za neke tradicionalne metode, kretanja se jednostavno procenjuju afnim modelima (Schultz i Stevenson, 1996). U (Protter et al., 2009, Takeda et al., 2009), usvajaju „nelokalne srednje vrednosti“ i „3D regresiju upravljačkog kernela“ za video super rezoluciju. Liu i Sun (2014) predložili su „Bajesov pristup“ istovremenoj proceni osnovnog kretanja, jezgra zamućenja i nivoa šuma za rekonstrukciju frejmova visoke rezolucije. U Ma et al. (2015), usvojena je metoda maksimizacije očekivanja za procenu jezgra zamućenja i rekonstrukciju frejmova visoke rezolucije. Međutim, ovi eksplicitni modeli video zapisa visoke rezolucije još uvek nisu adekvatni za uklapanje u razne scene u video zapisima.

Zbog velikog uspeha dubokog učenja u različitim oblastima (Zhang et al., 2021), sve je veće interesovanje za algoritme super rezolucije zasnovane na dubokom učenju. Ovi algoritmi zasnovani su na različitim arhitekturama dubokih neuronskih mreža, kao što: su konvolucione neuronske mreže (CNN, eng. *Convolutional Neural Network*), generativne adversarialne mreže (GAN, eng. *Generative Adversarial Network*) i rekurentne neuronske mreže (RNN,

eng. *Recurrent Neural Network*). Oni koriste veliki broj LR i HR video sekvenci za unos u neuronsku mrežu, potrebnih za poravnanje između okvira, ekstrakciju/fuziju karakteristika, a zatim za kreiranje sekvence za odgovarajući video niske rezolucije. Tok izvršenja (eng. *Pipeline*) većine VSR metoda uključuje jedan modul poravnanja, jedan modul za ekstrakciju i fuziju karakteristika i jedan modul za rekonstrukciju, kao što je prikazano na *Slici 1*.



Slika 1. Pipeline opšte namene metoda dubokog učenja za VSR zadatke.

Treba imati na umu da modul poravnjanja između okvira može biti zasnovan na tradicionalnim metodama ili na dubokom CNN, dok modul za ekstrakciju i fuziju karakteristika i modul za povećanje uzorkovanja obično koristi duboke CNN-ove. Isprekidana linija na *Slici 1* označava da je modul opcion.

Zbog sposobnosti nelinearnog učenja dubokih neuronskih mreža, metode zasnovane na dubokom učenju obično postižu dobre performanse na mnogim javnim benchmark skupovima podataka. Kao što je već rečeno, glavna razlika između VSR i ISR je u obradi informacija između frejmova.

2. Ocena kvaliteta videa

Za razliku od ISR koji ima za cilj da izvrši super rezoluciju jedne degradirane slike, VSR se bavi degradiranim video sekvencama i oporavljanjem do odgovarajuće HR video sekvence, koje treba da budu što sličnije realnim HR video sekvencama. Konkretno, VSR algoritam može koristiti slične tehnike kao ISR algoritam za obradu jednog okvira (prostorne informacije), međutim mora uzeti u obzir i relacije među frejmovima (vremenske informacije) da bi se obezbedila konzistentnost pokreta videa.

Kvalitet videa se najčešće izračunava pomoću maksimalnog odnosa signal-šum (PSNR, eng. *Peak Signal to Noise Ratio*) i indeksa strukturne sličnosti (SSIM, eng. *Structural Similarity*). Ovi indeksi mere razliku piksela i sličnost strukture između dve slike, respektivno. PSNR jednog SR frejma je definisan kao:

$$PSNR = 10 \log_{10} \left(\frac{L^2}{MSE} \right),$$

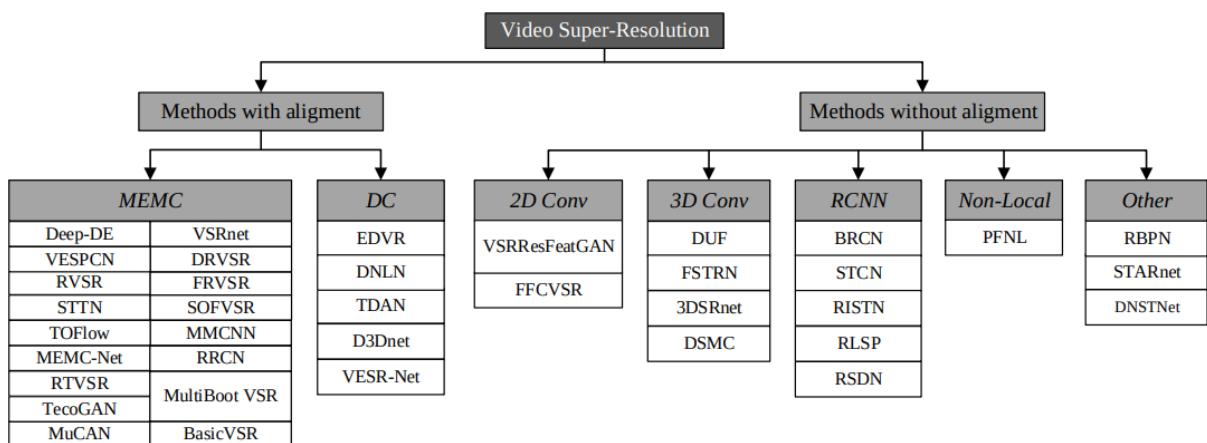
gde L predstavlja predstavlja maksimalni opseg vrednosti boje, što je obično 255, a MSE (eng. *Mean Square Error*) predstavlja srednju kvadratnu grešku.

3. Metode video super rezolucije

Kako su video snimci sačinjeni od pokretnih slika i zvuka, metode za video super-rezoluciju oslanjaju se na metode super-rezolucije jedne slike. Postoji mnoštvo metoda dubokog učenja namenjenih super rezoluciji slike: SRCNN (eng. *Super-Resolution using deep Convolutional Neural Networks*) (Dong et al., 2014), na osnovu koga je Kappeler predstavio metod video super-rezolucije pomoću konvolucionih neuronskih mreža (VSRnet) (Kappeler et al., 2016), FSRCNN (eng. *Fast Super-Resolution Convolutional Neural Networks*) (Dong et al., 2016), VDSR (Kim et al., 2016), ESPCN (eng. *Efficient Sub-Pixel Convolutional Neural Network*) (Shi et al., 2016), RDN (eng. *Residual Dense Network*) (Zhang et al., 2018), RCAN (eng. *Residual Channel Attention Network*) (Zhang et al., 2018b), ZSSR (eng. “Zero-Shot” Super-Resolution) (Shocher et al., 2018) i SRGAN (eng. *Super-Resolution using a Generative Adversarial Network*) (Ledig et al., 2017).

Nekoliko nedavnih studija o video super rezoluciji (e.g. Vang et al., 2019a, Jo et al., 2018, Tian et al., 2020), je pokazalo da korišćenje informacija sadržanih u frejmovima u velikoj meri utiče na performanse. Pravilna i adekvatna upotreba takvih informacija može poboljšati rezultate video super-rezolucije.

Postojeće metode se mogu podeliti u dve kategorije (*Slika 2*): metode sa poravnanjem i metode bez poravnanja, prema tome da li su video ramovi eksplisitno poravnati.



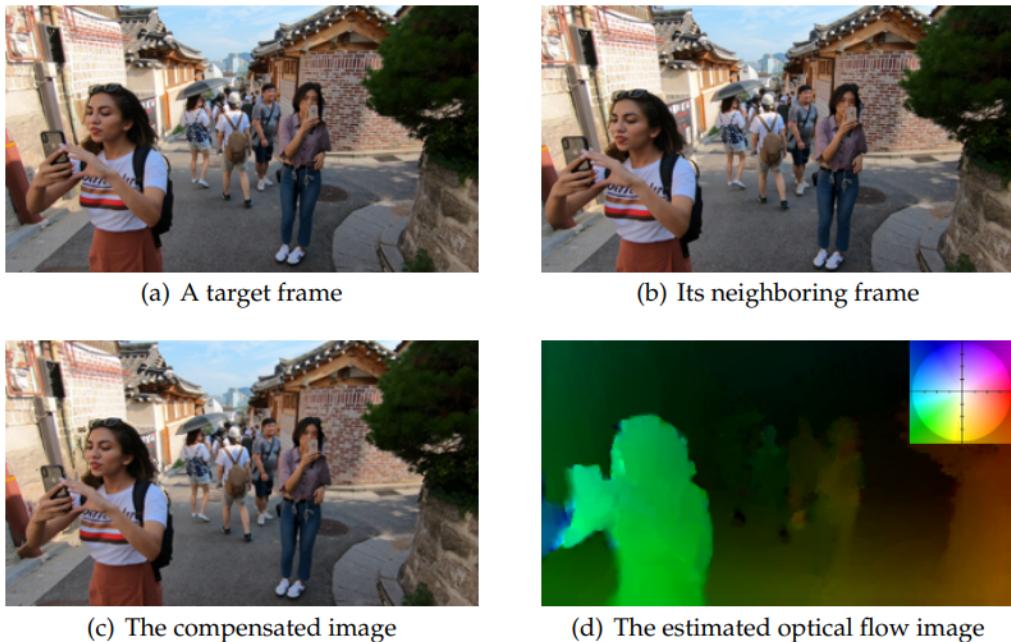
Slika 2. Podela metoda super rezolucije prema poravnjanju

3.1 Metode sa poravnanjem

Metode sa poravnanjem čine da se susedni okviri eksplisitno usklađuju sa ciljnim okvirom pomoću ekstrahovane informacije o kretanju pre naknadne rekonstrukcije. Ove metode uglavnom koriste procenu kretanja i kompenzaciju pokreta (MEMC, eng. *Motion Estimation, Motion Compensation*) ili deformabilnu konvoluciju, što su dve uobičajene tehnike za poravnavanje okvira.

3.1.1 Metode procene pokreta i kompenzacije

Većina metoda sa poravnanjem za video super-rezoluciju primenjuje procenu kretanja i kompenzaciju pokreta. Konkretno, svrha procene kretanja je da se izdvoji informacija o kretanju, dok se kompenzacija pokreta koristi za obavljanje operacije umotavanja (eng. *Wrapping*) između okvira, prema informacijama o kretanju između okvira, da bi se okviri poravnali jedan sa drugim. Većina tehnika za procenu kretanja se izvodi metodom optičkog toka (Dosovitskiy et al., 2015). Ovaj metod pokušava da izračuna kretanje između dva susedna okvira kroz njihove korelacije i varijacije u vremenskom domenu. Metode procene kretanja mogu se podeliti u dve kategorije: tradicionalne metode (npr. Lucas i Kanade, 1981 ili Drulea i Nedevschi, 2011) i metode dubokog učenja kao što su *FlowNet* (Dosovitskiy et al., 2015), *FlowNet 2.0* (Ilg et al., 2017) i *SpyNet* (Ranjan and Black, 2017). Primer procene i kompenzacije je prikazan na sledećoj slici.



Slika 3. Primer procene kretanja i kompenzacije. Različite boje predstavljaju različite pravce kretanja, a intenzitet boje je opseg kretanja.

3.2 Metode bez poravnjanja

Za razliku od metoda sa poravnanjem, metode bez poravnanja ne poravnavaju susedne okvire, već koriste prostorne ili prostorno-vremenske informacije za izdvajanje svojstava. Tehnike koje se koriste za početno izdvajanje karakteristika dele se u pet kategorija: metode 2D konvolucije (2D Conv, eng. *2D Convolution*), metode 3D konvolucije (3D Conv, eng. *3D Convolution*), rekurentna konvolucionna neuronska mreža (RCNN, eng. *Recurrent Convolutional Neural Network*), metode zasnovane na nelokalnoj mreži i druge metode.

3.2.1 2D konvolucione metode

Umesto operacija poravnjanja, kao što je kretanje procena i kompenzacija kretanja između okvira, ulazni okviri se direktno unose u 2D konvolucionu mrežu da bi se izvršila prostorna ekstrakcija karakteristika, fuzija i operacije super-rezolucije. Ovo može biti jednostavan pristup za rešavanje problema video super-rezolucije jer mreža sama uči informacije o korelaciji okvira.

3.2.2 3D konvolucione metode

3D konvolucijski modul (Tran et al., 2015, Ji et al., 2013) radi na prostorno-vremenskom domenu, u poređenju sa 2D konvolucijom, koja prostorne informacije koristi kroz klizno jezgro preko ulaznog okvira. Ovo je korisno za obradu video sekvene, gde se korelacija okvira postiže izdvajanjem vremenskih informacija. Većina 3D metoda ima relativno veću računsku složenost u poređenju sa 2D konvolucionim metodama, što ih ograničava kada je u pitanju super rezolucije videa u realnom vremenu.

3.2.3 Rekurentne konvolucione neuronske mreže (RCNNs)

Metode zasnovane na RCNN-u su pogodne za modeliranje prostorno-vremenskih informacija sadržanih u video snimcima, pošto mogu da mapiraju susedne okvire, i na taj način efikasno uspostavljaju dugoročnu zavisnost. Međutim, konvencionalne metode zasnovane na RCNN-u je teško obučiti, a javlja se i problem nestajanja gradijenta. Zato se može desiti da se ne obuhvati dugoročna zavisnost kada je dužina ulaznih sekvenci prevelika, što direktno utiče na performanse. Ova ograničenja se donekle mogu prevazići uz pomoć pamćenja karakteristika iz plićih slojeva (LSTM metode, eng. *Long Short-Term Memory*). Međutim, složen dizajn LSTM-a je faktor koji ograničava njihovu dubinu na hardveru, ograničavajući ih da modeluju veoma duge zavisnosti.

3.2.4 Ne-lokalne metode

Metoda koje nisu lokalno zasnovane takođe koriste i prostorne i vremenske informacije sadržane u video okvirima za super-rezoluciju. Ovaj metod zasniva se na ideji ne-lokalne neuronske mreže (Vang et al., 2018), koja je predložena za rad sa dugoročnim zavisnostima kod video klasifikacije. Ne-lokalne metode prevazilaze nedostatke konvolucionih i rekurentnih izračunavanja ograničenih na lokalno područje.

3.2.4 Druge metode

Metode u ovoj potkategoriji ne koriste početna izdvajanja obeležja (eng. *Features*), već

kombinuju više tehnika za super rezoluciju.

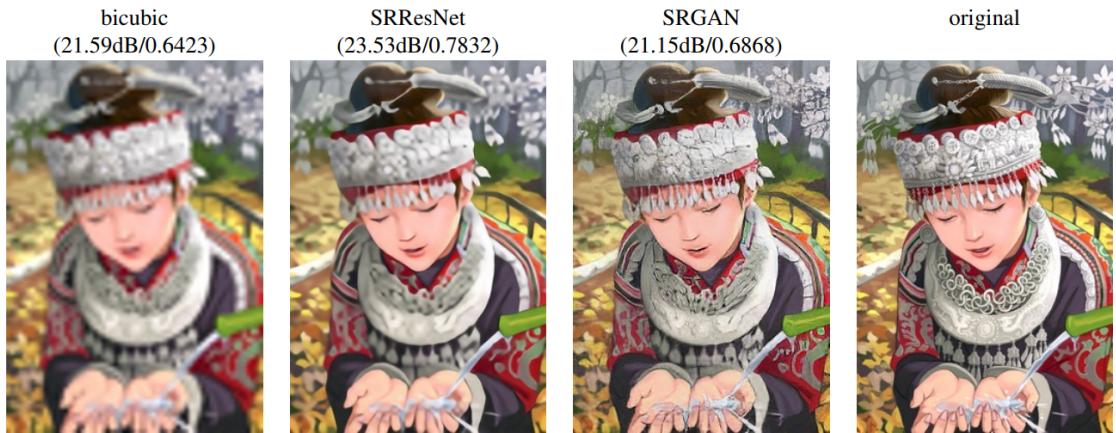
4. Realizacija sistema namenjenog super rezoluciji videa

Kao što je već rečeno, poslednjih godina predloženo je mnogo algoritama video super rezolucije, koji ostvaruju odlične rezultate. Ovo je ohrabrilo autore ovog izveštaja da izvrše implementaciju sopstvenog sistema po uzoru na neku od SOA (eng. *State Of the Art*) arhitektura. Primeri SOA arhitektura koje su uzete u razmatranje: „*Real-World Super-Resolution via Kernel Estimation and Noise Injection*“ [2], „*SwinIR: Image Restoration Using Swin Transformer*“ [3], „*Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data*“ [4], „*VRT: A Video Restoration Transformer*“ [5], „*BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond*“ [6]. Autori ovog izveštaja su proveli više nedelja pokušavajući da pokrenu i finaliziraju treniranje pomenutih arhitektura (i njihovih različitih implementacija), kako bi ih nakon toga specijalizovali za određeni zadatak, na primer: unapređenje kvaliteta arhivskog materijala (starih, crno–belih slika ili videa). Ovaj pristup nije urođio plodom na dostupnom hardveru (*NVIDIA GeForce RTX 3050 Ti Laptop GPU, 11th gen. i7 processor*), kao ni na *Colab* platformi. Hardverski zahtevi datih mreža su u originalu, nažalost, preveliki. Autori najčešće nisu bili u mogućnosti da pokrenu treniranje mreža čije su modele (naivno) uprostili smanjivši broj različitih slojeva višestruko. Došli su do zaključka da su mreže sve kompleksnije i da je sve manje šansi da se one obuče u “kućnoj” varijanti.

Iz tog razloga, realizacija sistema podrazumevala je korišćenje pretreniranog modela. Nakon što su više nedelja proveli proučavajući razne arhitekture, odlučili su se za korišćenje “Real-ESRGAN” (eng. *Real Enhanced Super-Resolution using a Generative Adversarial Network*) [7]. Ovaj model predstavlja proširenje ESRGAN (eng. *Enhanced Super-Resolution using a Generative Adversarial Network*) modela, koji je zasnovan na modelu za super-rezoluciju slike SRGAN (eng. *Super-Resolution using a Generative Adversarial Network*). O svakom od ovih modela biće više reči u nastavku.

5. SRGAN

Problem super rezolucije je posebno izražen za visoke faktore povećanja (eng. *Upscale Factor*), kod kojih nedostaju detalji tekstura na rekonstruisanim slikama. Optimizacija SR algoritama obično podrazumeva minimizovanje srednje vrednosti kvadratne greška između oporavljene i orginalne HR slike. Ovo je zgodno jer minimizovanje MSE maksimizuje vršni odnos između signala i šuma (PSNR), što je uobičajena mera koja se koristi za procenu i poređenje SR algoritama. Međutim, sposobnosti MSE (i PSNR) da se odrede perceptivno relevantne razlike, npr. visoki detalji teksture, veoma su ograničene jer su definisane na osnovu razlike u pikselima. Ovo je prikazano na *Slici 4*, gde viši PSNR nužno ne znači i bolji SR rezultat.



Slika 4. S leva na desno: bikubična interpolacija, duboka rezidualna mreža optimizovana za MSE, duboka rezidualna generativna adversarialna mreža optimizovana za gubitak osetljiv na ljudsku percepciju, originalna HR slika. Odgovarajući PSNR i SSIM su prikazani u zagradama. [4× povećanje]

U nastavku prikazana je SRGAN, mreža koja koristi duboku rezidualnu mrežu (*ResNet*) sa skip-vezom (eng. *Skip-connection*) i odstupa od MSE-a kao jedinog cilja optimizacije. U odnosu na prethodne radove, gubitak percepcije definisan je koristeći mape karakteristika (eng. *Feature maps*) visokog nivoa VGG [8] mreže, u kombinaciji sa diskriminatorom. Primer fotorealistične slike, nad kojom je izvršena super rezolucija sa faktorom povećanja od 4×, prikazan je na slici 5.



Slika 5. Super razrešena slika (levo) se gotovo ne razlikuje od originalne (desno). [4× povećanje]

5.1 Povezani radovi

5.1.1 Super rezolucija slike

U okviru ovog poglavlja, fokus je na super rezoluciji jedne slike (SISR, eng. *Single Image Super-Resolution*). Metode zasnovane na predviđanju bile su među prvim metodama koje su se koristile za SISR. Iako ovi pristupi, npr. linearni, bikubični ili „Lanczos filtriranje“, mogu biti veoma brzi, oni previše pojednostavljaju problem SISR i obično daju rešenja sa previše

glatkim teksturama. Moćniji pristupi oslanjaju se na podatke u obuci, za koje postoje LR i HR parovi.

SR algoritmi zasnovani na konvolucionim neronskim mrežama (CNN), pokazali su odlične performanse. Sposobnost mreže da nauči filtere za uvećanje (eng. upscaling), direktno utiče na performanse u smislu tačnosti i brzine.

5.1.2 Dizajn konvolucionih neuronskih mreža

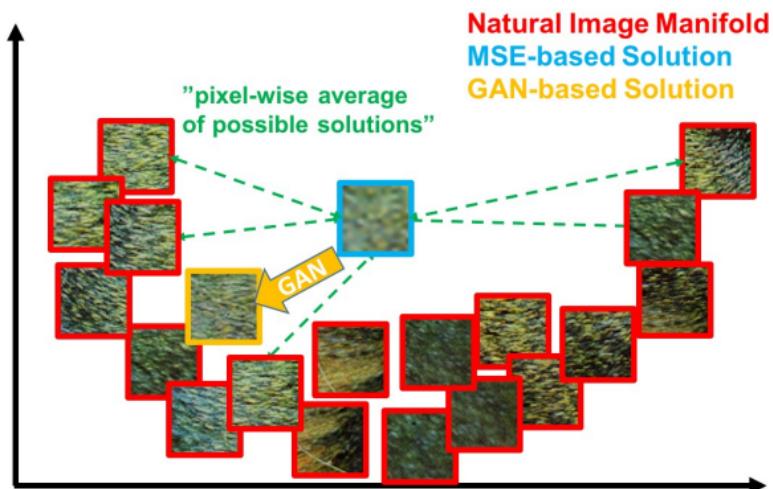
Uspeh rešenja mnogih problema u kompjuterskom vidu direktno zavisi od arhitekture konvolucionih neuronskih mreža. Pokazalo se da je dublje mrežne arhitekture teže trenirati, ali mogu u značajnoj meri povećavaju tačnost mreže jer omogućavaju preslikavanja veoma visoke složenosti. Za efikasno obučavanje dubljih mrežnih arhitektura često se koristi batchnormalization za suzbijanje internih promena ko-varijansi. Pokazalo se da dublje mrežne arhitekture takođe povećavaju performanse za SISR. Još jedan moćan izbor dizajna koji olakšava obuka dubokih CNN-a je nedavno uveden koncept rezidualnih blokova i skip-veza. Preskočne veze olakšavaju mrežnu arhitekturu modeliranja mapiranja identiteta, koje je trivijalno po prirodi, međutim, potencijalno netrivijalno za predstavljanje sa konvolucionim jezgrima.

U kontekstu SISR-a, takođe se pokazalo da je učenje uvećanja filtera korisno u smislu tačnosti i brzine.

5.1.3 Funkcije gubitka

Funkcije gubitka (eng. *Loss functions*), kao što je MSE, loše su u obradi detalja visoke frekvencije, kao što je tekstura. MSE podstiče pronalaženje proseka u pikselima, pa su rešenja tipično preterano glatka i stoga imaju loš perceptivni kvalitet.

Na *Slici 6* ilustrovan je problem minimiziranja MSE, gde su prikazana višestruka potencijalna rešenja da bi se stvorila glatka rekonstrukcija. Rešenje zasnovano na MSE izgleda „previše glatko“ dovodi do perceptualno loših rezultata. Na *Slici 4* je dat prikaz rezultata dobijenih PSNR funkcije gubitka.



Slika 6. Ilustracija zakrpa (eng. patches) sa prirodnih slika (crvene) i super-razrešene zakrpe dobijene sa MSE (plava) i GAN (narandžasta).

U Mathieu et al. [8] i Denton et al. [8], autori predlažu rešenje ovog problema primenom generativne adversarialne mreže (GAN). Dosovitskiy i Brox [8] koriste funkcije gubitaka zasnovane na euklidskim rastojanjima izračunatim u prostoru obeležja neuronske mreže u kombinaciji sa adversarnom obukom. Pokazano je da predloženi gubitak omogućava generisanje vizuelno superiornije slike i može se koristiti za rešavanje loše postavljenog inverznog problema dekodiranja nelinearnih reprezentacija obeležja. Slično ovom radu, Johnson et al. [8] i Bruna et al. [8] predlažu korišćenje funkcija ekstrahovanih iz unapred obučene VGG mreže umesto mere greške u pikselima niskog nivoa. Konkretno, autori formulišu funkciju gubitka na osnovu euklidske udaljenosti između mapa obeležja izdvojenih iz VGG19 mreže [8], kako bi se ostvarili ubedljivi rezultati za obe super-rezolucije.

5.2 Adversarialna mreža

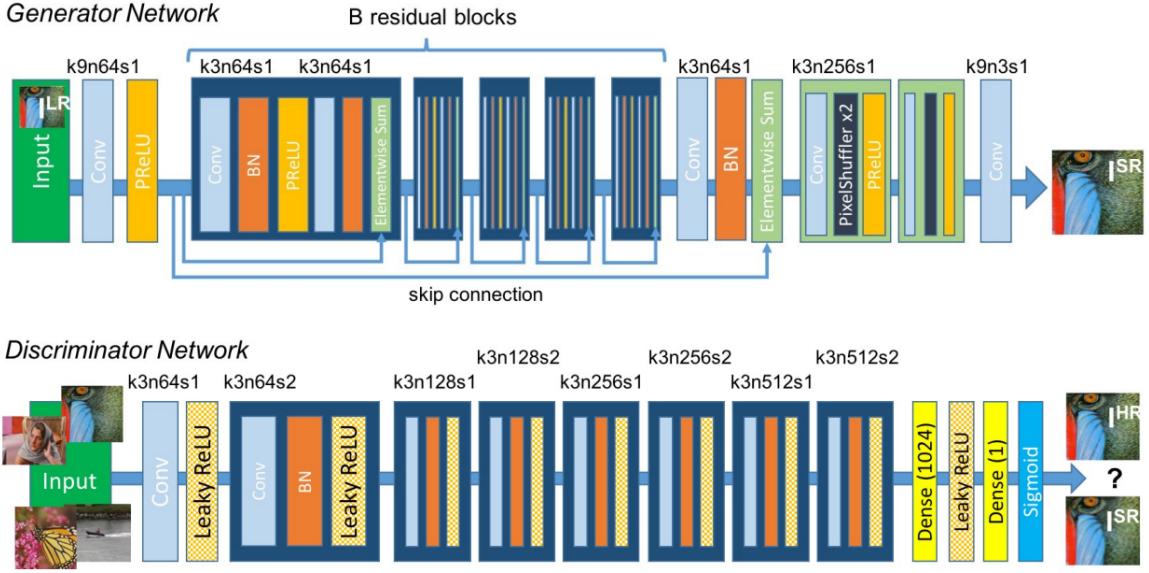
SISR ima za cilj da izvrši super-rezoluciju nad ulaznom slikom niske rezolucije, kako bi se dobila slika visoke rezolucije. Slike visoke rezolucije su dostupne samo u toku treninga. Super-rezolucija same slike u velikoj meri zavisi od GAN mreže koja se koristi.

Da bi se prevazišao adversarialni min-max problem, definisana je sledeća formula:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \\ \mathbb{E}_{I^{LR} \sim p_G(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))]$$

Opšta ideja iza ove formulacije je da ona dozvoljava da se obuči generativni model G sa ciljem da se zavara diferencibilni diskriminator D koji je obučen da razlikuje super-razrešene slike od stvarnih slika. Sa ovakvim pristupom, generator može naučiti da kreira rešenja koja su veoma slična stvarnim slikama i stoga ih je teško klasifikovati od strane D . Ovo se razlikuje od SR rešenja dobijenih minimiziranjem funkcije gubitaka vezane za vrednosti piksela, kao što je MSE.

U osnovi veoma duboke generatorske mreže G , koja je ilustrovana na narednoj *Slici 4*, nalazi se B rezidualnih blokova sa identičnim rasporedom. Konkretno, koriste se dva konvolucionia sloja sa malim jezgrima 3×3 i 64 mape karakteristika praćene slojevima za normalizaciju serije [32] i ParametricReLU [28] kao aktivaciona funkcija. Rezolucija ulazne slike se povećava pomoću dva obučena konvolucionia sloja podpiksela.



Slika 7. Arhitektura generatorske i diskriminatorske mreže sa odgovarajućom veličinom kernela (k), brojem mapa obeležja (n) i korakom (s) naznačenim za svaki konvolucijski sloj.

Za razlikovanje stvarnih HR slika od generisanih SR uzoraka koristi se diskriminatorska mreža. Arhitektura mreže je prikazana na *Slici 7*. U toku razvoja arhitekture, praćene su arhitektonske smernice koje je predložio Radford et al. [8], koristi se LeakyReLU aktivaciona funkcija ($\alpha = 0,2$) i izbegava max-pooling. Diskriminatorska mreža je obučena za rešavanje *min-max* problema opisanog u prethodnoj jednačini. Ona se sastoji od osam konvolucionih slojeva kod kojih raste broj 3×3 filter kernela. Broj ovih krenela raste s faktorom 2, sa 64 na 512 jezgara kao u VGG mreži. Konvolucije sa korakom se koriste za smanjenje rezolucije slike svaki put kada je broj karakteristika udvostručen. Nakon dobijenih 512 mapa obeležja, slede dva gusta sloja (eng. *Dense layers*), i konačno sigmoidna aktivaciona funkcija za dobijanje verovatnoće za klasifikaciju uzorka

5.3 Perceptualna funkcija gubitaka

Definicija funkcije perceptivnog gubitka l^{SR} je kritična za performanse mreže generatora. Dok se l^{SR} obično modeluje na osnovu MSE, dizajnirana je funkcija gubitka koja procenjuje rešenje oslanjajući se na perceptivno relevantne karakteristike. Ova funkcija predstavlja poboljšanje u odnosu na Johnson et al. [8] i Bruna et al. [8]. Perceptivni gubitak se formuliše kao zbir težina gubitka sadržaja (l_X^{SR}) i adversarialne komponente gubitka:

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}$$

perceptual loss (for VGG based content losses)

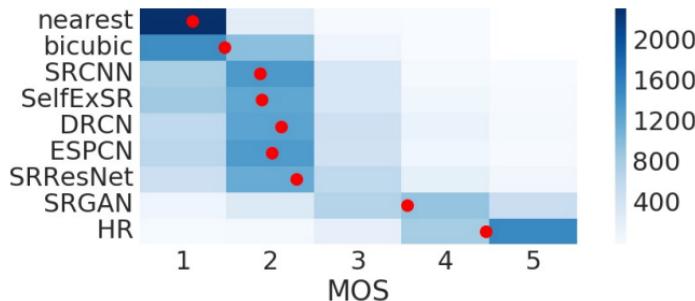
5.4 Poređenje sa drugim modelima

Poređenje se izvodi nad tri široko korišćena skupa podataka za referentne vrednosti *Set5* [8], *Set14* [8] i *BSD100* [8]. Na eksperimentima gde se vrši skaliranje slika niske rezolucije,

sa faktorom $4\times$, kako bi se doobile slike visoke rezolucije, pokazano je da standardne kvantitativne mere, kao što su PSNR i SSIM, ne uspevaju da precizno prikažu kvalitet slike prema ljudskom vizuelnom sistemu. Rezultati poređenja dati su u nastavku:

	SRResNet-		SRGAN-		
	MSE	VGG22	MSE	VGG22	
PSNR	32.05	30.51	30.64	29.84	29.40
SSIM	0.9019	0.8803	0.8701	0.8468	0.8472
MOS	3.37	3.46	3.77	3.78	3.58
Set14					
PSNR	28.49	27.19	26.92	26.44	26.02
SSIM	0.8184	0.7807	0.7611	0.7518	0.7397
MOS	2.98	3.15*	3.43	3.57	3.72*

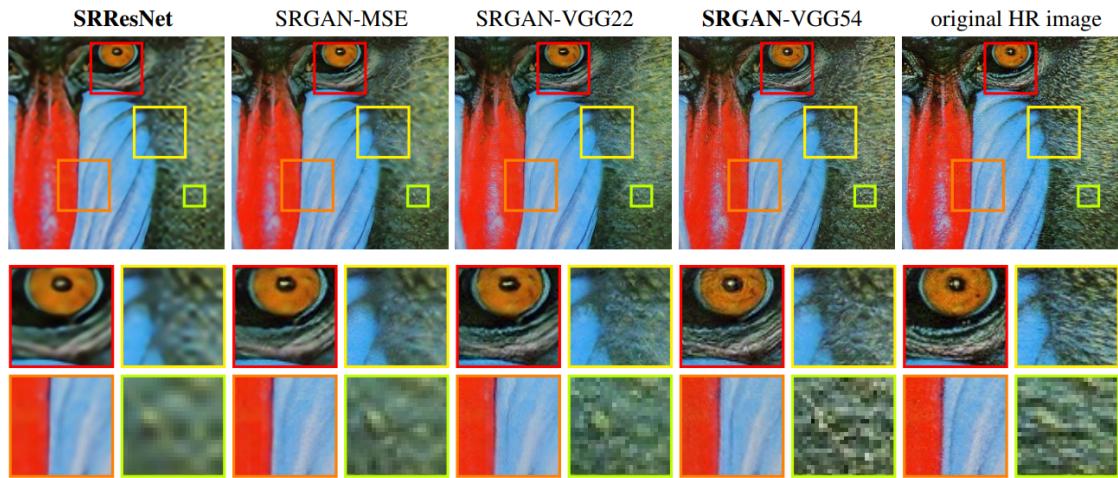
Tabela 1. Performanse SRResNet i SRGAN modela nad Set5 i Set14 skupom podataka.



Slika 8. Raspodela MOS rezultata označena je bojama nad skupom podataka BSD100

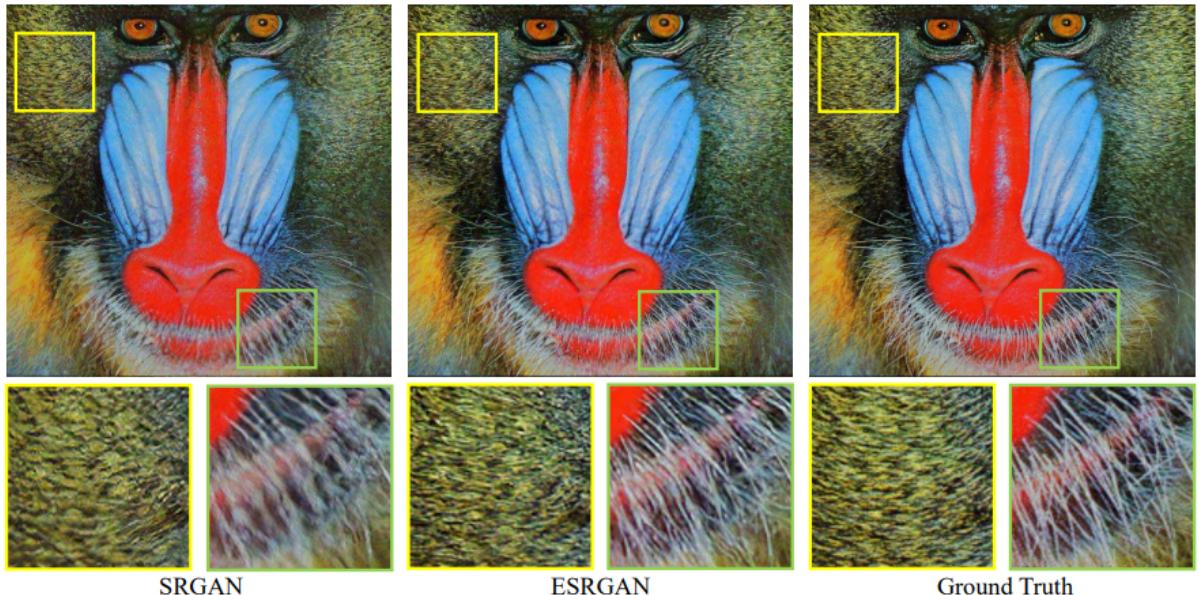
Set5	nearest	bicubic	SRCCNN	SelfExSR	DRCN	ESPCN	SRResNet	SRGAN	HR
PSNR	26.26	28.43	30.07	30.33	31.52	30.76	32.05	29.40	∞
SSIM	0.7552	0.8211	0.8627	0.872	0.8938	0.8784	0.9019	0.8472	1
MOS	1.28	1.97	2.57	2.65	3.26	2.89	3.37	3.58	4.32
Set14									
PSNR	24.64	25.99	27.18	27.45	28.02	27.66	28.49	26.02	∞
SSIM	0.7100	0.7486	0.7861	0.7972	0.8074	0.8004	0.8184	0.7397	1
MOS	1.20	1.80	2.26	2.34	2.84	2.52	2.98	3.72	4.32
BSD100									
PSNR	25.02	25.94	26.68	26.83	27.21	27.02	27.58	25.16	∞
SSIM	0.6606	0.6935	0.7291	0.7387	0.7493	0.7442	0.7620	0.6688	1
MOS	1.11	1.47	1.87	1.89	2.12	2.01	2.29	3.56	4.46

Tabela 2. Prikazuje sumarizovane rezultati za modele: SRResNet, SRGAN, Nearest Neighbor, SRCNN [9], SelfExSR [31], DRCN [34], ESPCN [48], SRResNet i SRGAN-VGG54, nad sva tri skupa podataka.



Slika 8. Vizuelni rezultati za SRResNet i SRGAN modele [4x uvećanje]

6. ESRGAN



Slika 9. Poređenje stvarne slike i rezultata super rezolucije za SRGAN i ESRGAN sa uvećanjem od 4x. ESRGAN nadmašuje SRGAN u oštini i detaljima.

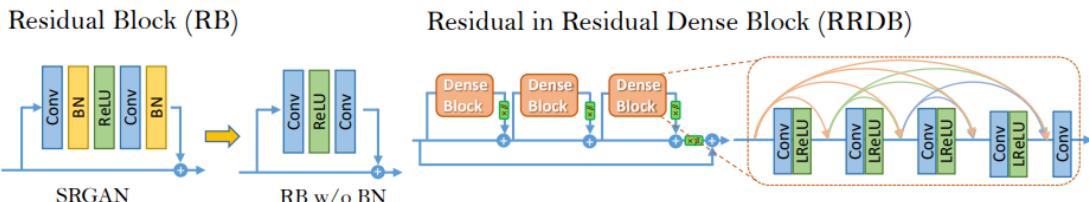
Sa prethodne slike se može videti da postoji razlika između SRGAN rezultata i stvarne slike. ESRGAN predstavlja unapređen SRGAN model sa poboljšanjima koja se odnose na strukturu mreže uvođenjem RDDB (eng. *Residual-in-Residual Dense Block*), koji je većeg kapaciteta i lakši za obuku, što je opisano u [9]. Uklonjeni su BN (eng. *Batch Normalization*) slojevi, umesto kojih se koristi rezidualno skaliranje i manja inicijalizacija da bi se olakšala obuka veoma duboke mreže. Poboljšan je diskriminator korišćenjem RaGAN (eng. *Relativistic average GAN*), koja uči da proceni „da li je jedna slika realističnija od druge“ nego „da li je jedna slika prava ili lažna“. Eksperimentalnim putem je pokazano da ova poboljšanja pomažu generatoru da povrati realističnije detalje teksture. Još jedno poboljšanje odnosi se na gubitak percepcije korišćenjem VGG funkcija pre aktivacije, umesto nakon

aktivacije kao u SRGAN-u. Empirijski je pokazano da prilagođeni perceptivni gubitak daje oštije ivice i vizuelno prijatnije rezultate.

6.1 Arhitektura mreže

U cilju daljeg poboljšanja kvaliteta slike SRGAN-a, napravljene su dve modifikacije u strukturi generatora G :

- 1) ukoljeni su svi BN slojevi;
- 2) zamenjen je originalni osnovni blok sa predloženim RRDB blokom, koji kombinuje rezidualnu mrežu na više nivoa i gustim vezama (eng. *Dense connections*) kao što je prikazano na *Slici 10*.



Slika 10. Levo: Uklonjeni su BN slojevi u rezidualnom bloku SRGAN-a. Desno: RRDB blok se koristi u dubljem modelu, β je rezidualni parametar skaliranja.

Pokazalo se da uklanjanje BN slojeva povećava performanse i smanjuje složenost računanja u različitim zadacima orijentisanim na PSNR, uključujući SR i uklanjanje zamućenja (eng. *Deblurring*). BN slojevi normalizuju karakteristike koristeći srednju vrednost i varijansu u seriji (eng. *Batch*) tokom obuke i koriste procenjenu srednju vrednost i varijansu celog skupa podataka za obuku tokom testiranja. Kada se statistika obuke i testiranja skupova podataka dosta razlikuju, BN slojevi imaju tendenciju da uvode artefakte i ograničavaju sposobnost generalizacije. Empirijski je primećeno da je veća verovatnoća da će BN slojevi uneti artefakte kada je mreža dublja i obučena korišćenjem GAN-ova. Ovi artefakti se povremeno pojavljuju među iteracijama i različitim postavkama, narušavajući stabilan učinak tokom treninga, zato uklanjanje BN slojeva doprinosi stabilnom treningu i konzistentnom učinku. Dodatno se poboljšava sposobnost generalizacije i smanjuje složenost računanja i korišćenja memorije.

Zadržan je dizajn arhitekture visokog nivoa SRGAN-a, a koristi se novi osnovni blok, odnosno RRDB kao što je prikazano na *Slici 10*. Na osnovu zapažanja da više slojeva i veza može da poveća performanse, predloženi RRDB blok koristi dublju i složeniju strukturu od originalnog rezidualnog bloka u SRGAN-u. Konkretno, kao što je prikazano na *Slici 10*, predloženi RRDB blok ima „residual-in-residual“ strukturu, gde se rezidualno učenje koristi u različitim nivoima. Opisani RRDB blok koristi „guste“ (eng. Dense) blokove u glavnoj putanji, tako da kapacitet mreže postaje veći korišćenjem gustih veza.

Pored poboljšane arhitekture, koristi se i nekoliko tehnika da bi se olakšala obuka veoma dubokih mreža:

- 1) Skaliranje residualnih vrednosti konstantom koja uzima vrednost između 0 i 1, kako bi se sprečila nestabilnost;

2) uzimanje manjih vrednosti prilikom inicijalizacije, jer je empirijski potvrđeno da je rezidualnu arhitekturu lakše obučiti kada je početna varijansa parametra manja.

6.2 Relativistički diskriminator

Pored poboljšane strukture generatora, unapređen je i diskriminator na osnovu relativističkog GAN-a. Za razliku od standardnog diskriminatora D u SRGAN-u, koji procenjuje verovatnoću da je jedna ulazna slika x realna i prirodna, relativistički diskriminator pokušava da predvidi verovatnoću da je prava slika x_r je relativno realističnija od lažne x_f , kao što je prikazano na *Slici 11*.

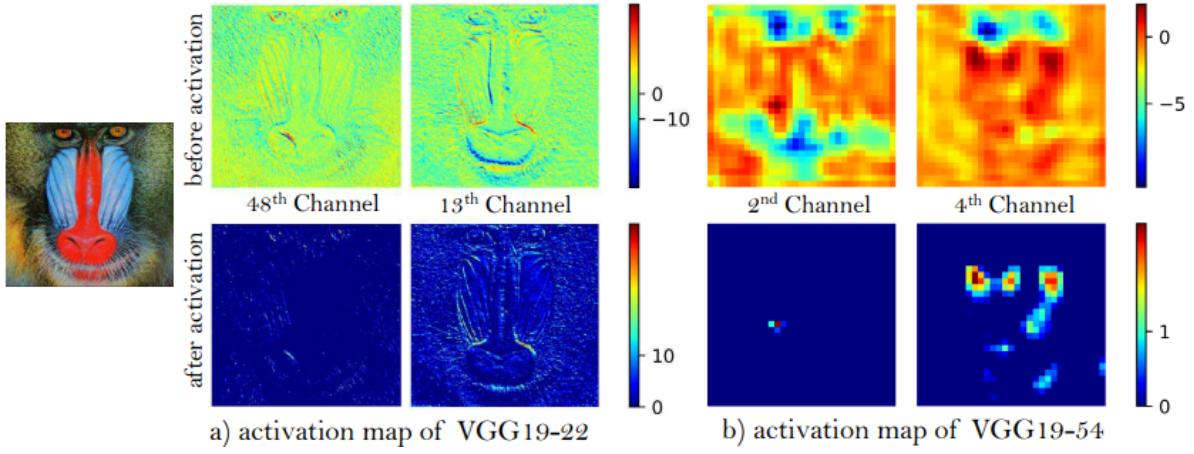
The diagram illustrates the difference between a Standard GAN and a Relativistic GAN. On the left, under 'a) Standard GAN', it shows two equations: $D(x_r) = \sigma(C(\text{Real})) \rightarrow 1$ labeled 'Real?' and $D(x_f) = \sigma(C(\text{Fake})) \rightarrow 0$ labeled 'Fake?'. An orange arrow points from this section to the right section. On the right, under 'b) Relativistic GAN', it shows two equations: $D_{Ra}(x_r, x_f) = \sigma(C(\text{Real}) - \mathbb{E}[C(\text{Fake})]) \rightarrow 1$ labeled 'More realistic than fake data?' and $D_{Ra}(x_f, x_r) = \sigma(C(\text{Fake}) - \mathbb{E}[C(\text{Real})]) \rightarrow 0$ labeled 'Less realistic than real data?'. Below each section is a small image of a face, with the 'Real' image being more detailed and the 'Fake' image being more blurry.

Slika 11. Razlika između standardnog diskriminatora i relativističkog diskriminatora

6.3 Perceptualni gubitak

Razvijen je efikasniji perceptivni gubitak *Lpercep*, ograničavanjem karakteristika pre aktivacije, a ne posle aktivacije kao što se praktikuje u SRGAN-u.

Na osnovu ideje da su bliži perceptivnoj sličnosti, Johnson et al. [9] predlaže gubitak percepcije, koji je proširen u SRGAN-u. Perceptualni gubitak je prethodno definisan na aktivacionim slojevima unapred obučene duboke mreže, gde je rastojanje između dve aktivirane karakteristike minimizirano. Protivno konvenciji, predloženo je da se svojstva koriste pre slojeva aktivacije, da bi se prevazišla ova dva nedostatka originalnog dizajna. Prvo, aktivirana svojstva su veoma retka, posebno nakon veoma duboke mreže, kao što je prikazano na *Slici 12*. Na primer, prosečan procenat aktiviranih neurona za sliku „Babun“ posle VGG19-54 sloja je samo 11,17%. Oskudna aktivacija dovodi do lošijeg učinka. Drugo, korišćenje svojstva posle aktivacije takođe uzrokuje nedoslednu rekonstruisanu osvetljenost u poređenju sa istinitom slikom.



Slika 12. Reprezentativne mape karakteristika pre i posle aktivacije za sliku 'babun'. Kako mreža ide dublje, većina svojstva nakon aktivacije postaje neaktivna, dok svojstva pre aktivacije sadrže više informacija.

U suprotnosti sa uobičajeno korišćenim gubitkom percepcije koji usvaja VGG mreža obučena za klasifikaciju slika, razvijen je prikladniji perceptivni gubitak za SR – MINC gubitak. Zasnovan je na fino podešenoj VGG mreži za prepoznavanje materijala [3], koja se fokusira na teksture, a ne na objekat. Iako je dobitak od indeks percepcije, koji donosi MINC gubitak, marginalan, gubitak percepcije koji se fokusira na teksturu je ključan za SR.

6.4 Mrežna interpolacija

Da bi se uklonio šum u metodama zasnovanim na GAN-u, a da se pritom očuva dobar perceptivni kvalitet, predložena je fleksibilna i efikasna strategija – mrežna interpolacija. Konkretno, prvo je obučena PSNR orijentisana mreža GPSNR, a zatim je finim podešavanjem GAN mreže dobijena mreža GGAN. Interpolirani su svi odgovarajući parametri ove dve mreže, da bi se dobio interpolirani model G_{INTERP} . Predložena mrežna interpolacija ima dve prednosti. Prvo, interpolirani model je u stanju da proizvede značajne rezultate za bilo koji α , bez uvođenje artefakata. Drugo, može se kontinuirano uravnotežiti perceptivni kvalitet i tačnost, bez preobuke modela.

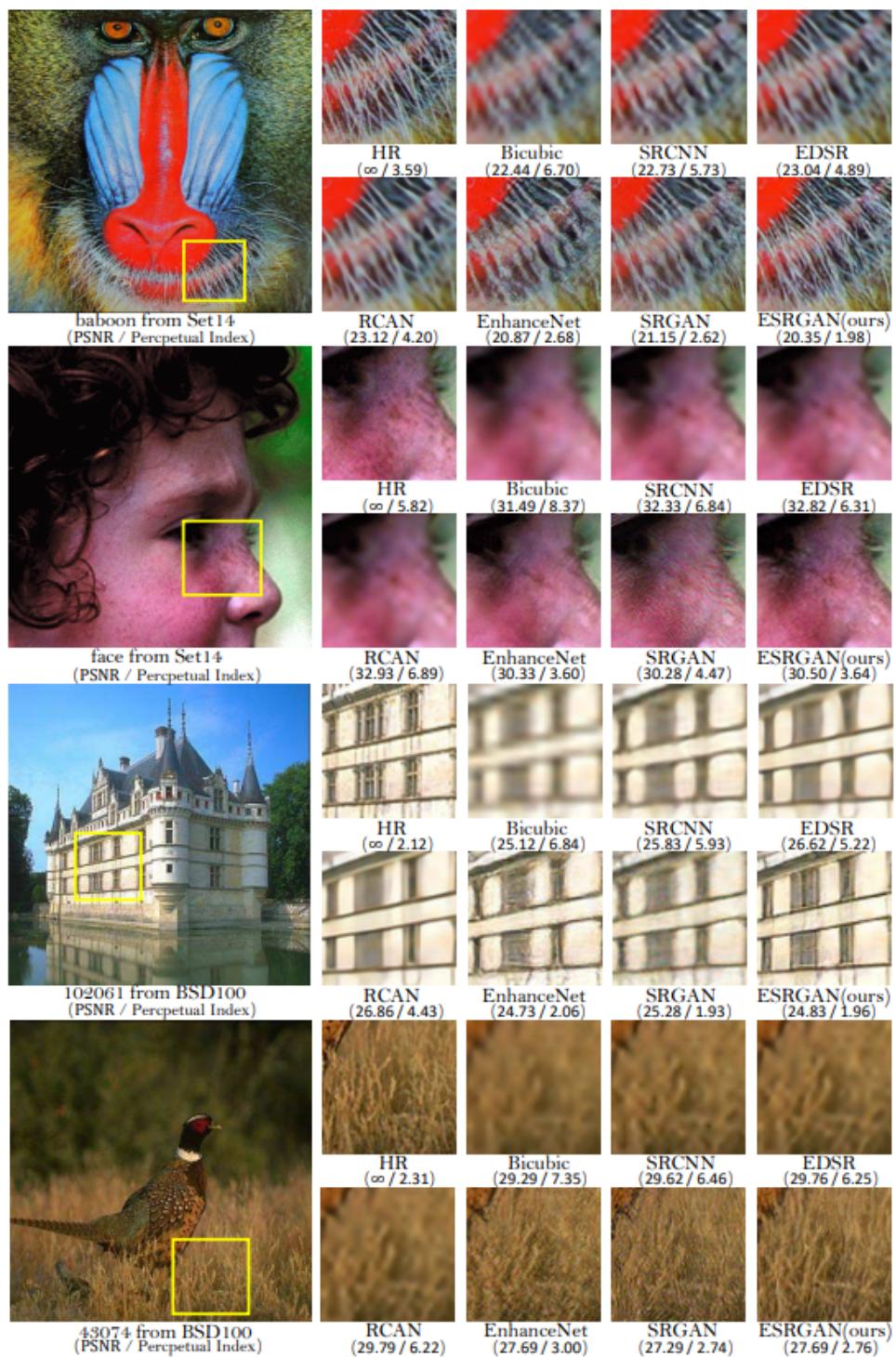
Takođe, istražene su alternativne metode za balansiranje PSNR-orientisanih efekata i metoda zasnovanih na GAN-u. Na primer, može se direktno interpolirati njihov izlaz slika (piksel po piksel) umesto mrežnih parametara. Međutim, takav pristup ne uspeva da postigne dobar kompromis između šuma i zamućenja, tj. interpolirana slika je ili previše mutna ili sadrži previše šuma sa artefaktima. Drugi pristup je da se podese težine gubitka sadržaja i suprotstavljenog gubitka, međutim ovaj pristup zahteva podešavanje težina i fino podešavanje mreže, pa je stoga preskupo da bi se postigla kontinuirana kontrola stila slike.

6.5 Rezultati ESRGAN modela

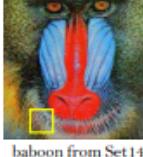
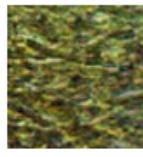
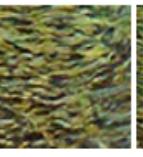
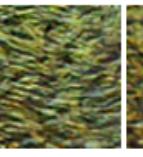
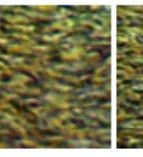
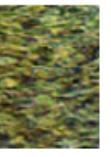
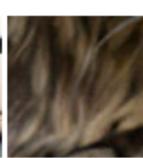
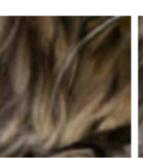
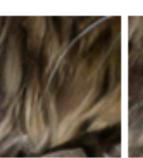
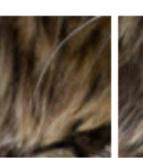
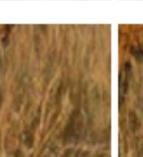
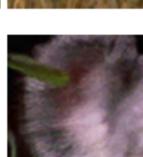
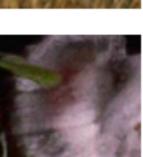
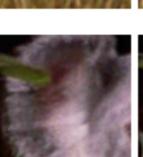
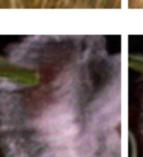
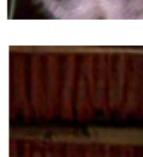
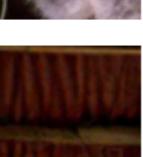
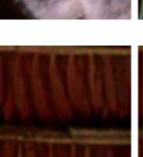
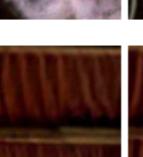
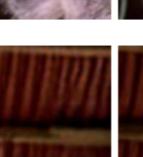
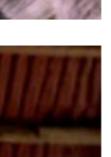
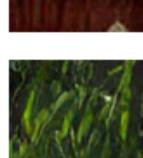
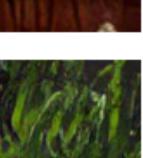
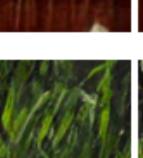
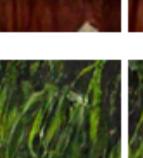
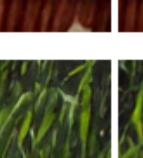
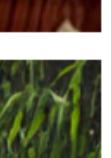
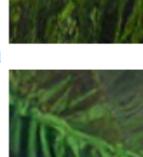
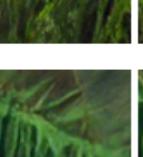
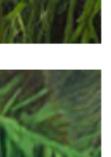
Izvršeno je poređenje opisanog modela na nekoliko skupova javnih referentnih podataka sa najsavremenijim metodama orijentisanim na PSNR, uključujući SRCNN, EDSR i RCAN, a takođe i sa pristupima vodenim percepcijom, uključujući SRGAN i EnhanceNet.

Kvalitativni rezultati su prikazani na *Slici 13*, sa koje se može videti da predloženi ESRGAN model ima bolji učinak u oštrini i detaljima. Na primer, ESRGAN može da proizvode oštije i prirodne brkove pavijana i teksturu trave (*Slika 13*) od metoda orijentisanih na PSNR, koje imaju tendenciju da generišu zamućene rezultate, i od prethodnih metoda zasnovanih na GAN-u, čije su teksture neprirodne i sadrže šum. ESRGAN je u stanju da generiše detaljnije teksture na zgradi (*Slika 13*), dok druge metode ili ne uspevaju da proizvedu dovoljno detalja (SRGAN) ili dodaju neželjene teksture (EnhanceNet). Štaviše, prethodne metode zasnovane na GAN-u ponekad uvode nepoželjne artefakte, npr. SRGAN dodaje bore licu. ESRGAN model se oslobađa ovih artefakata i daje prirodne rezultate.

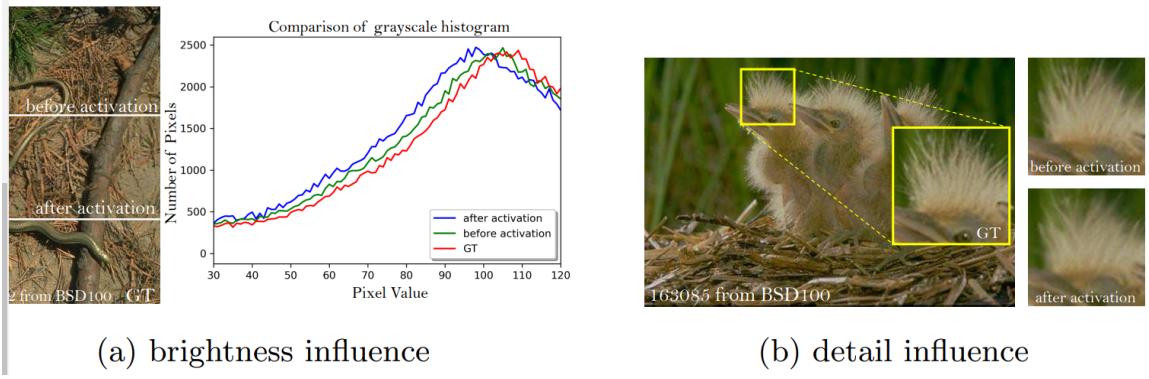
Na *Slici 14* prikazani su rezultati ESRGAN mreže kada se svojstva koriste pre i posle aktivacije. Za razliku od SRGAN-a koji tvrdi da je modele sa sve većom dubinom teže obučavati, model koji koristi ESRGAN pokazuje superiorne performanse, a uz to ga je i lako obučavati, bez obzira na veliku dubinu, zahvaljujući pomenutim poboljšanjima, posebno predloženim RRDB bez BN slojeva.



Slika 13. Kvalitativni rezultati ESRGAN-a. ESRGAN u odnosu na SRGAN proizvodi prirodneje teksture, na primer: životinjsko krzno, artefakti na licu, građevinsku strukturu i teksturu trave sa manjom količinom neželjenih artefakata

1 st	2 nd	3 rd	4 th	5 th	6 th	7 th
BN?	✓	✗	✗	✗	✗	✗
Activation?	After	After	Before	Before	Before	Before
GAN?	Standard GAN	Standard GAN	Standard GAN	RaGAN	RaGAN	RaGAN
Deeper with RRDB?	✗	✗	✗	✗	✓	✓
More data?	✗	✗	✗	✗	✗	✓
 baboon from Set14						
 baboon from Set14						
 39 from PIRM self_val						
 43074 from BSD100						
 69015 from BSD100						
 6 from PIRM self_val						
 20 from PIRM self_val						
 208001 from BSD100						

Slika 14. Rezultati ESRGAN mreže nakon uvedenih poboljšanja



Slika 15. Poređenje rezultata kada se svojstva koriste pre i nakon aktivacije

U nastavku je izvršeno poređenje efekta mrežne interpolacije i strategije interpolacije slike. Primjenjuje se jednostavna linearna interpolacija. Interpolacioni parametar α se bira u opsegu vrednosti od 0 do 1 sa korakom od 0,2.

Kao što je prikazano na *Slici 16*, čista metoda zasnovana na GAN-u proizvodi oštре ivice i bogatije teksture, ali sa nepoželjnim artefaktima, dok čista PSNR orijentisana metoda daje mutne slike u stilu crtanog filma. Korišćenjem mrežne interpolacije smanjuju se neprijatni artefakti dok se teksture održavaju. Nasuprot tome, interpolacija slike ne uspeva da efikasno ukloni ove artefakte.



Slika 16. Poređenje između mrežne interpolacije i interpolacije slike

7. Real-ESRGAN

Iako je učinjeno mnogo pokušaja u super-rezoluciji da se restauriraju slike niske rezolucije koje sadrže složene degradacije, rešenje opšte degradirane slike iz stvarnog sveta je još uvek daleko. U nastavku je proširen moćni ESRGAN model kako bi se bolje simulirale složene degradacije u stvarnom svetu. Opisan je način na koji se prevazilaze artefakti u procesu sinteze. Dodatno, koristi se U-Net diskriminator [10] sa spektralnom normalizacijom da bi se povećala sposobnost diskriminatora i stabilizovala dinamika treninga. Poređenjem dobijenih rezultata može se uočiti da su Real-ESRGAN performanse superiorne nad raznim stvarnim skupovima podataka u odnosu na prethodne rade.

7.1 Degradacija slika

Degradacije u stvarnom svetu su obično previše složene da bi se modelovale jednostavnim kombinacijama višestrukih degradacija. Prave složene degradacije obično potiču od komplikovanih kombinacija različitih procesa degradacije, kao što su sistemi kamera za snimanje slika, uređivanje slika i internet prenos. Na primer, kada se kreira fotografija mobilnim telefonom, fotografija može sadržati nekoliko degradacija, kao što su zamućenje kamere, šum senzora, izoštravanje artefakta i „JPEG kompresija“. Zatim se nad fotografijom vrše dodatne izmene korišćenjem raznih alata, nakon čega se fotografija otprema na društvene mreže, koje vrše dalju kompresiju. Navedeni proces postaje komplikovaniji kada se slika više puta deli na Internetu.

Ovo je uticalo na proširenje klasičnog modela "prvog reda" degradacije na model degradacije „visokog reda“ za degradacije u stvarnom svetu, tj. degradacije su modelovane sa nekoliko ponovljenih procesa degradacije, gde svaki proces predstavlja klasičan model degradacije. Predložena je strategija nasumične promene da bi se sintetizovale praktičnije degradacije. Međutim, to još uvek uključuje fiksni broj procesa degradacije, i ostaje nejasno da li su sve izmešane degradacije korisne ili ne. Umesto toga, modeliranje degradacija visokog reda je fleksibilnije jer pokušava da oponaša stvarni proces generisanja degradacija. U proces sinteze uključuju se „sink filteri“ (eng. *sinc filters*) da bi se simulirali uobičajeni artefakti zvonjave i prekoračenja (eng. *Ringing and overshoot artifacts*).

Pošto je prostor za degradaciju mnogo veći od ESRGAN-a, obuka takođe postaje izazovna. Konkretno:

1) diskriminator zahteva veću moć razlikovanja. Odnosno, da u toku obuke razlikuje stvarne od složenih rezultata, dok gradijent povratne informacije diskriminatora treba da bude tačniji za lokalno poboljšanje detalja. Stoga, VGG diskriminator u ESRGAN-u je poboljšan korišćenjem U-Net dizajna.

2) U-Net struktura i komplikovane degradacije takođe povećavaju nestabilnost treninga, pa se iz tog razloga koristi spektralna normalizacija (SN, eng. *Spectral Normalization*) regularizacije za stabilizaciju dinamike treninga.

Zahvaljujući pomenutim poboljšanjima, Real-ESRGAN se može lako trenirati, tako da se ostvaruje dobar balans lokalnog poboljšanja detalja i suzbijanja artefakata.

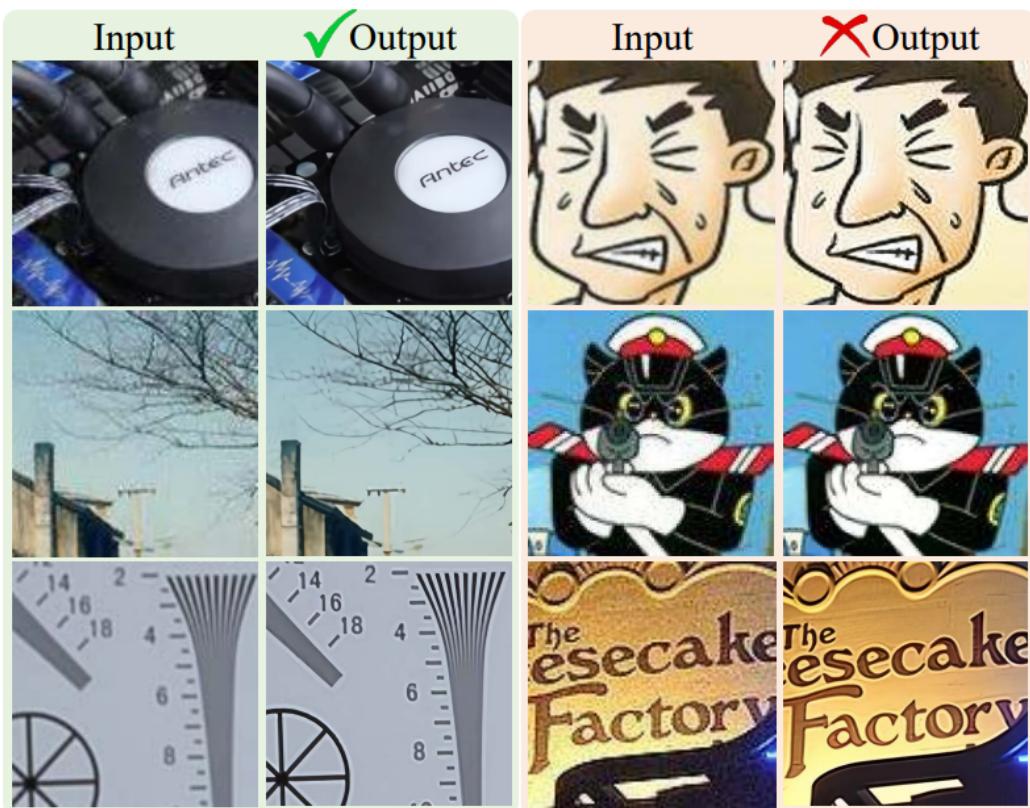
7.2 Model degradacije visokog reda

Ako bi se koristio klasični model degradacije, opisan u [10], da bi se sintetizovali parovi za obuku, obučeni model bi zaista mogao da radi sa pravim uzorcima. Međutim, on još uvek ne može rešiti neke komplikovane degradacije u stvarnom svetu, posebno nepoznati šum i složene artefakte (*Slika 17*), zato što postoji značajna razlika između sintetičkih slika niske rezolucije i realističnih degradiranih slika.

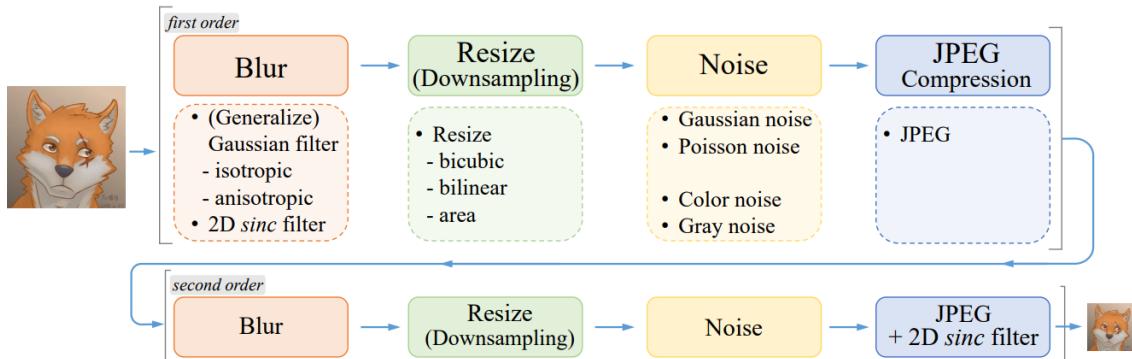
Klasični model degradacije uključuje samo fiksni broj osnovnih degradacija, što se može posmatrati kao modeliranje prvog reda. Međutim, procesi degradacija u stvarnom svetu prilično su raznovrsni i obično se sastoje od niza parametara, uključujući: sistem snimanja slika kamera, softvera za uređivanje slika, internet prenosa itd. Na primer, ako je potrebno primeniti super-rezoluciju nad slikom niskog kvaliteta preuzetom sa interneta, njena osnovna degradacija uključuje kombinaciju različitih procesa degradacije. Konkretno, originalna slika može biti snimljena mobilnim telefonom pre mnogo godina, što neminovno sadrži degradacije kao što su zamućenje kamere, šum senzora, niska rezolucija i „JPEG kompresija“. Slika je zatim uređivana operacijama izoštravanja i promene veličine, što je dovelo do artefakta prekoračenja i zamućenja. Nakon toga, slika je postavljena na neku od društvenih mreža, što uvodi dalju kompresiju i nepredvidivi šum. Digitalni prenos takođe unosi artefakte, što ovaj proces čini komplikovanijim kada se slika nekoliko puta prenosi internetom.

Ovako komplikovan proces degradacije ne bi mogao biti rešen korišćenjem klasičnog modela prvog reda. Stoga, predložen je model degradacije visokog reda. Model „n-tog reda“ uključuje n ponovljenih procesa degradacije, gde svaki proces degradacije usvaja klasični model degradacije sa istom procedurom, ali različitim hiperparametrima. Treba imati na umu da se „visoki red“ (eng. *High-order*) razlikuje od onog koji se koristi u matematičkim funkcijama. To se uglavnom odnosi na „vreme implementacije“ iste operacije. Strategija slučajnog mešanja takođe može uključivati ponovljeni proces degradacije (npr. dvostruko zamućenje ili „JPEG kompresiju“). Međutim, bitno je istaknuti da je proces degradacije visokog reda ključan, što ukazuje da nisu sve izmešane degradacije neophodne. Da bi se rezolucija slike održala u razumnom opsegu, operacija smanjenja uzorkovanja se zamenjuje operacijom nasumične promene veličine. Empirijski, usvojen je proces degradacije drugog reda, jer može da reši većinu stvarnih slučajeva uz zadržavanje jednostavnosti. Na *Slici 18* je dat prikaz procesa sintetičkog generisanja podataka.

Vredi napomenuti da poboljšani proces degradacije visokog reda nije savršen i da ne može da pokrije ceo prostor degradacije u stvarnom svetu. Umesto toga, on samo proširuje mogućnosti degradacije prethodnih SR metoda kroz modifikovanje procesa sinteze podataka. Primeri ovih ograničenja se mogu videti na *Slici 23*.



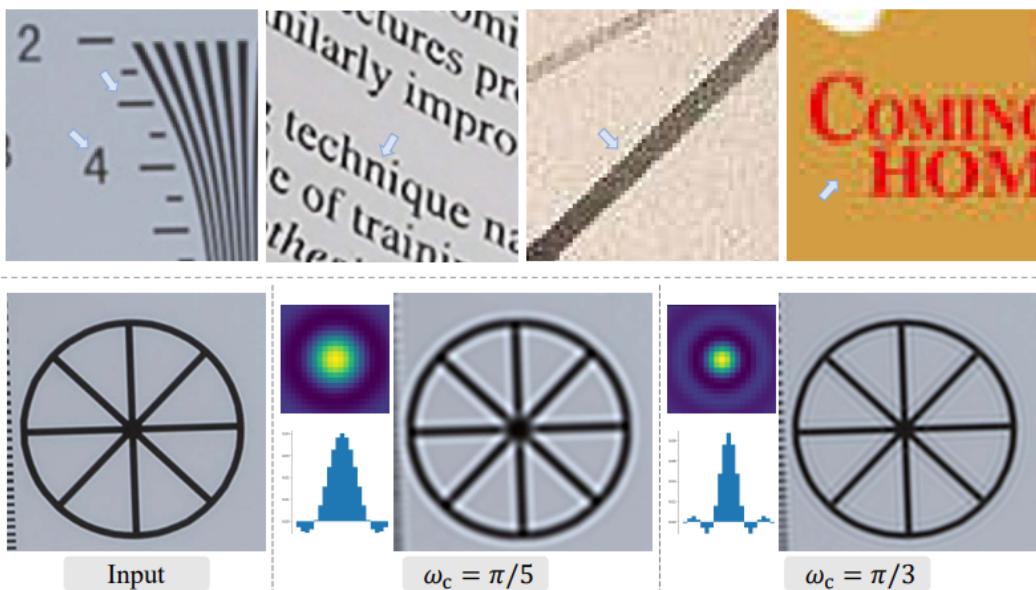
Slika 17. Modeli obućeni sintetičkim podacima, korišćenjem klasičnog modela degradacije. Ovaj model je u stanju da reši neke stvarne uzorce (levo), ali pojačava šum ili uvodi artefakte za složene slike iz stvarnog sveta (desno).



Slika 18. Pregled čistog sintetičkog generisanja podataka usvojenog u Real-ESRGAN-u.

7.3 Artefakti zvonjave i prekoračenja

Artefakti koji zvone često se pojavljuju kao lažne ivice u blizini oštrih prelaza na slici. Oni vizuelno izgledaju kao „duhovi“ blizu ivica. Artefakti prekoračenja obično se kombinuju sa artefaktima zvonjave, koji se manifestuju kao povećan skok na prelazu ivice. Glavni uzrok ovih artefakata su signali ograničeni opsezima bez visoke frekvencije. Ovi artefakti su uobičajeni i veoma često proizvedeni algoritmom za izoštrevanje, JPEG kompresijom itd. Primer ovih artefakta dat je na *Slici 19*.



Slika 19. Gore: Pravi uzorci koji pate od artefakta koji zvone i artefakta prekoračenja. Dole: Primeri sink kernela (veličina jezgra 21) i odgovarajuće filtrirane slike.

Koristi se sink filter, idealizovani filter koji isključuje visoke frekvencije, da bi se sintetizovala zvonjava i prekoračenje artefakta za trening parove. Kernel sink filtera može biti izražen kao:

$$k(i, j) = \frac{\omega_c}{2\pi\sqrt{i^2 + j^2}} J_1(\omega_c\sqrt{i^2 + j^2}),$$

gde su (i, j) koordinate kernela, W_c je frekvenca odsecanja, a J_1 je „Beselova funkcija“ prvog reda, prve vrste. U donjem delu Slike 19 su dati prikazi sink filtera sa različitim graničnim frekvencijama i njihove odgovarajuće filtrirane slike. Može se primetiti da sink filteri dobro sintetišu artefakte zvonjave i artefakte prekoračenja. Ovi artefakti su vizuelno slični onima koji se javljaju u prva dva stvarna uzorka, kao u gornjem delu Slike 19.

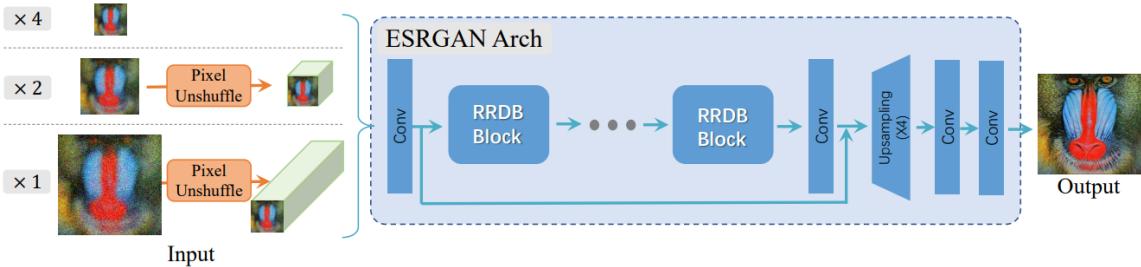
Sink filteri se koriste na dva mesta: u procesu zamućenja i u poslednjem koraku sinteze. Poslednji sink filter i JPEG kompresija se nasumično smenuju da bi se pokrio veći prostor za degradaciju, na primer kao što su slike koje su prvo previše izoštrene (sa artefaktima prekoračenja), a zatim se nad njima vrši JPEG kompresija, ili slike nad kojima se prvo vrši JPEG kompresija, a zatim operacija izoštravanja.

7.4 Mreže i obuka

7.4.1 ESRGAN generator

Real-ESRGAN koristi isti generator kao ESRGAN mreža, i sadrži nekoliko rezidualnih gustih blokova (RRDB), prikazan je na Slici 20. Na istoj slici se može uočiti da za faktor razmere $\times 2$ i $\times 1$, prvo se koristi operacija poništavanja piksela da bi se smanjila prostorna veličina i izvršila preraspodela informacija u dimenziji kanala. Takođe, proširena je originalna $\times 4$ ESRGAN arhitektura da bi se ostvarila super-rezolucija sa faktorom skale od $\times 2$ i $\times 1$. Pošto je ESRGAN “teška” mreža, prvo se koristi metoda preuređenja piksela (eng.

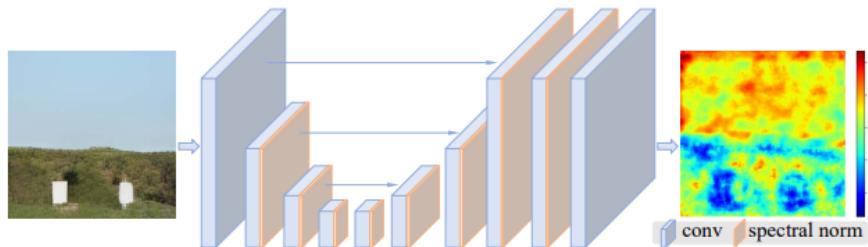
Pixel-unshuffle, inverzna operacija u odnosu na „*Pixel-shuffle*“) da bi se smanjila prostorna veličina i povećala veličina kanala pre unosa ulaza u glavnu ESRGAN arhitekturu. Dakle, većina proračuna se vrši u manjem prostoru rezolucije, što može smanjiti GPU memoriju i potrošnju računskih resursa.



Slika 20. Real-ESRGAN usvaja istu mrežu generatora kao i ESRGAN.

7.4.2 U-Net diskriminator sa spektralnom normalizacijom (SN)

Kako Real-ESRGAN ima za cilj da reši mnogo veći prostor degradacije od ESRGAN-a, originalni dizajn diskriminatora u ESRGAN-u više nije prikladan. Konkretno, diskriminator u Real-ESRGAN-u zahteva veću diskriminatorsku moć za kompleksnije rezultate. Umesto da diskriminiše globalne stilove, on treba da proizvode tačne povratne informacije o gradijentu za lokalne teksture. Inspirisan [39, 50], diskriminator u VGG stilu u ESRGAN-u je poboljšan na U-Net dizajn sa skip-vezama (Slika 21). UNet daje vrednost tačnosti za svaki piksel i može da obezbedi generatoru detaljne povratne informacije za svaki piksel.



Slika 21. Arhitektura U-Net diskriminatora sa spektralnom normalizacijom

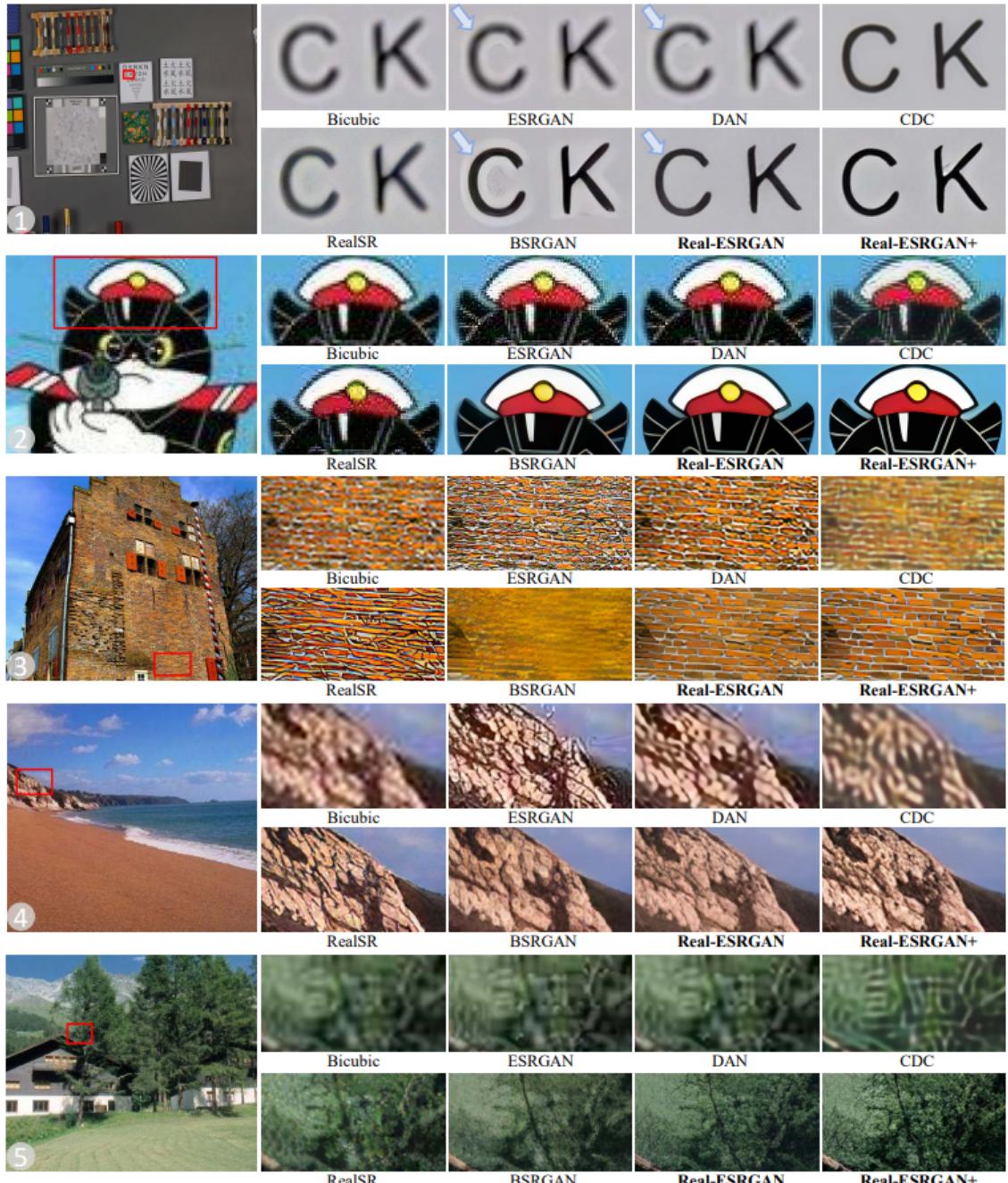
U-Net struktura i kompleksne degradacije takođe povećavaju nestabilnost treninga. Koristi se regularizacija spektralne normalizacije da se stabilizuje dinamika treninga. Štaviše, može se primetiti da je spektralna normalizacija takođe korisna za ublažavanje oštih artefakata koje uvodi GAN obuka. Sa tim prilagođavanjima, lako se može obučiti Real-ESRGAN mreža, uz postizanje poboljšanja lokalnih detalja i suzbijanja artefakata.

7.4.3 Proces obuke

Je podeljen u dve faze, prvo se obučava model orijentisan na PSNR sa gubitkom $L1$. Dobijeni model je nazvan Real-ESRNet. Zatim koristi se obučeni PSNR orijentisani model, kao inicijalizacija generatora, i trenira Real-ESRGAN kombinacijom $L1$ gubitka, perceptivnim gubitkom i GAN gubitkom.

7.5 Rezultati Real-ESRGAN modela

Izvršeno je poređenje Real-ESRGAN modela sa nekoliko najsavremenijih metoda, uključujući ESRGAN, DAN, CDC, RealSR i BSRGAN. Poređenje je izvršeno nad nekoliko skupova podataka za testiranje koji sadrže slike iz stvarnog sveta, uključujući: RealSR, DRealSR, OST300, DPED, ADE20K validacija i Internet slike, dok su rezultati prikazani na *Slici 22*.



Slika 22. Kvalitativna poređenja na nekoliko reprezentativnih uzoraka iz stvarnog sveta sa faktorom skale povećanja uzorkovanja od 4×.

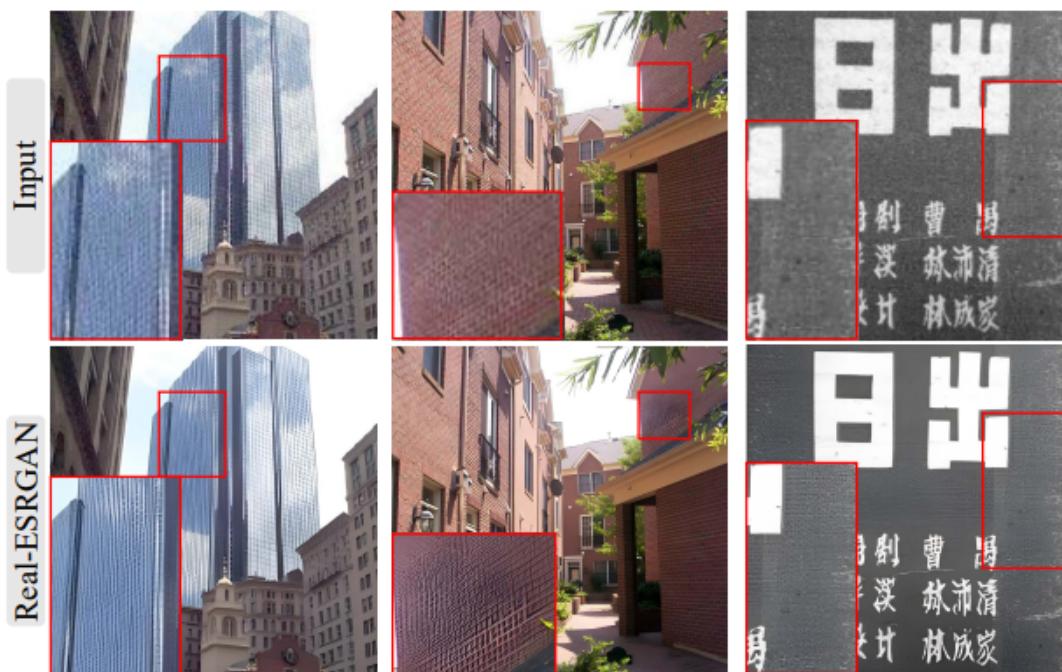
Real-ESRGAN nadmašuje prethodne pristupe u uklanjanju artefakata i restauraciji detalja

teksture. Real-ESRGAN+ (obučen sa izoštrenim slikama) može dodatno povećati vizuelnu oštrinu. Druge metode možda neće uspeti da uklone prekoračenje (prvi uzorak) i komplikovane artefakte (drugi uzorak), ili neće uspeti da vrate prirodne i realistične teksture za različite scene (3., 4., 5. uzorak).

Na *Slici 22* se može uočiti da Real-ESRGAN nadmašuje prethodne pristupe u uklanjanju artefakata i vraćanju detalja tekture. Real-ESRGAN+ (obučen sa izoštrenim slikama) može dodatno povećati vizuelnu oštrinu. Konkretno, prvi uzorak sadrži artefakte prekoračenja — bele ivice oko slova. Direktno povećanje uzorkovanja će neizbežno pojačati te artefakte (npr. DAN i BSRGAN). Real-ESRGAN uzima uobičajene artefakte u razmatranje i simulira ih pomoću sink filtera, čime se efikasno uklanjaju artefakti zvonjave i artefakti prekoračenja. Drugi uzorak sadrži nepoznate i komplikovane degradacije. Većina algoritama ih ne može efikasno eliminisati, dok bi Real-ESRGAN obučen sa procesima degradacije drugog reda mogao. Real-ESRGAN je takođe sposoban za vraćanje realističnijih tekstura (npr. cigla, planina i tekture drveta) za uzorce iz stvarnog sveta, dok druge metode ili ne uspevaju da uklone degradacije ili dodaju neprirodne teksture (npr. RealSR i BSRGAN).

7.6 Ograničenja

Iako Real-ESRGAN može da restaurira većinu slika iz stvarnog sveta, on i dalje ima neka ograničenja.



Slika 23. Ograničenja: 1) uvrnute linije; 2) neprijatni artefakti izazvani GAN obukom; 3) nepoznate i vandistributivne degradacije.

Kao što je prikazano na *Slici 23*:

- 1) neke restaurirane slike (naročito slike građevina i scena u zatvorenom) imaju uvrnute linije zbog problema sa aliasingom (eng. *aliasing*);

- 2) GAN obuka uvodi nepoželjne artefakte na nekim uzorcima;
 - 3) Real-ESRGAN nije mogao da ukloni komplikovane degradacije u stvarnom svetu.
- Može se dogoditi da se ovi artefakti čak i pojačaju.

Pomenuti nedostatci imaju veliki uticaj na praktičnu primenu Real-ESRGAN-a.

8. Zaključak

U ovom radu opisan je Real-ESRGAN, kao i modeli na kojima je on zasnovan. Za svaki od modela opisana je arhitektura, kao i poboljšanja koja on uvodi. Real-ESRGAN predstavlja “*State Of The Art*” metodu koja daje bolje rezultate od svih do sada poznatih metoda. Ova metoda je ustanju da izvrši super-rezoluciju slika iz stvarnog sveta, uklanjanjem artefakta i kreiranjem sadržaja visokog kvaliteta. Iako su rezultati za većinu slika i više nego dobri, još uvek postoji razlika između stvarnih i generisanih slika. Daljim razvojem Real-ESRGAN modela, i modela sličnih njemu, problem super-rezolucije bi mogao biti rešen.

Literatura

- [1] Video Super-Resolution Based on Deep Learning: A Comprehensive Survey -
<https://arxiv.org/pdf/2007.12928.pdf/>
- [2] Real-World Super-Resolution via Kernel Estimation and Noise Injection -
<https://paperswithcode.com/paper/real-world-super-resolution-via-kernel>
- [3] SwinIR: Image Restoration Using Swin Transformer -
<https://paperswithcode.com/paper/swinir-image-restoration-using-swin>
- [4] Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data - <https://paperswithcode.com/paper/real-esrgan-training-real-world-blind-super>
- [5] VRT: A Video Restoration Transformer -
<https://paperswithcode.com/paper/vrt-a-video-restoration-transformer>
- [6] BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond - <https://paperswithcode.com/paper/basicvsr-the-search-for-essential-components>
- [7] Real-ESRGAN - <https://github.com/xinntao/Real-ESRGAN/releases/>
- [8] Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network - <https://arxiv.org/pdf/1609.04802.pdf>
- [9] ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks -
https://openaccess.thecvf.com/content_ECCVW_2018/papers/11133/Wang_ESRGAN_Enhanced_Super-Resolution_Generative_Adversarial_Networks_ECCVW_2018_paper.pdf
- [10] Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data -
https://openaccess.thecvf.com/content/ICCV2021W/AIM/papers/Wang_Real-ESRGAN_Training_Real-World_Blind_Super-Resolution_With_Pure_Synthetic_Data_ICCVW_2021_paper.pdf