

Technická univerzita v Košiciach
Fakulta elektrotechniky a informatiky

**Modelovanie airglowu pomocou hlbokého
učenia**

Bakalárska práca

2022

Michal Bencúr

Technická univerzita v Košiciach
Fakulta elektrotechniky a informatiky

**Modelovanie airglowu pomocou hlbokého
učenia**

Bakalárska práca

Študijný program: Hospodárska informatika
Študijný odbor: Informatika
Školiace pracovisko: Katedra kybernetiky a umelej inteligencie (KKUI)
Školiteľ: doc. Ing. Peter Butka, PhD.
Konzultant: RNDr. Šimon Mackovjak, PhD.
Konzultant: Ing. Viera Maslej Krešňáková, PhD.

Košice 2022

Michal Bencúr

Abstrakt v SJ

V rozmedzí medzi zemskou atmosférou a vesmírom existuje veľké množstvo javov. Jedným z nich je aj žiarenie airglow. Cieľom tejto práce je vytvorenie modelu airglow-u pomocou hlbokého učenia z dát NASA misie GOLD. Práca poskytuje základné teoretické charakteristiky airglow-u, misie GOLD, hlbokého učenia a vybraného algoritmu pre vytvorenie modelu. Zaoberá sa taktiež vytvorením grafickej verzie dát misie GOLD, návrhom a tvorením modelu v jazyku Python. Výsledky predikcií modelu sú na záver porovnané s pôvodnými dátami pomocou vybraných metrík.

Kľúčové slová

airglow, hlboké učenie, misia GOLD, vesmírny výskum, predikcia

Abstrakt v AJ

There are many phenomena between the Earth's atmosphere and the space. One of them is airglow radiation. The aim of this thesis is to create an airglow model using deep learning from NASA's GOLD mission data. The work provides basic theoretical characteristics of airglow, GOLD mission, deep learning, and selected algorithm for model creation. This thesis also deals with the creation of a graphical version of GOLD mission data, design, and creation of a model in Python. Finally, the results of the model predictions are compared with the original data using selected metrics.

Kľúčové slová v AJ

airglow, deep learning, mission GOLD, space research, prediction

TECHNICKÁ UNIVERZITA V KOŠICIACH
FAKULTA ELEKTROTECHNIKY A INFORMATIKY
Katedra kybernetiky a umelej inteligencie

ZADANIE
BAKALÁRSKEJ PRÁCE

Študijný odbor: **Informatika**
Študijný program: **Hospodárska informatika**

Názov práce:

Modelovanie airglowu pomocou hlbokého učenia

Airglow modeling by deep learning

Študent: **Michal Bencúr**
Školiteľ: **doc. Ing. Peter Butka, PhD.**
Školiace pracovisko: **Katedra kybernetiky a umelej inteligencie**
Konzultant práce: **Ing. Viera Maslej Krešňáková, PhD., RNDr. Šimon Mackovjak, PhD.**
Pracovisko konzultanta: **Ústav experimentálnej fyziky SAV**

Pokyny na vypracovanie bakalárskej práce:

1. Podať teoretický prehľad problematiky modelovania airglowu pomocou hlbokého učenia.
2. Analyzovať a predspracovať dostupnú množinu satelitných dát zobrazujúcich airglow.
3. Navrhnuť, realizovať a otestovať postupy pre predikciu airglowu pomocou zvolených metód hlbokého učenia.
4. Vypracovať dokumentáciu podľa pokynov katedry a vedúceho práce.

Jazyk, v ktorom sa práca vypracuje: slovenský
Termín pre odovzdanie práce: 27.05.2022
Dátum zadania bakalárskej práce: 29.10.2021



prof. Ing. Liberios Vokorokos, PhD.
dekan fakulty

Čestné vyhlásenie

Vyhlasujem, že som bakalársku prácu vypracoval samostatne s použitím uvedenej odbornej literatúry.

Košice 3. 6. 2022

.....

Vlastnoručný podpis

Podakovanie

Týmto by som chcel poďakovať doc. Ing. Petrovi Butkovi, PhD., RNDr. Šimonovi Mackovjakovi, Phd. a Ing. Viere Maslej Krešňákovkej, PhD. za možnosť pracovať na tejto bakalárskej práci, za ich ochotu, rady a taktiež za ich čas, ktorý mi venovali pri tvorbe tejto práce.

Obsah

Úvod	1
1 Doména skúmania	2
1.1 Airglow	2
1.2 NASA misia GOLD	4
1.3 Pozorovania misie	9
2 Hlboké učenie	12
2.1 Strojové učenie	12
2.2 Generatívne adverzné siete	14
2.2.1 Pix2Pix model	15
2.3 Použité knižnice	17
2.4 Metriky pre porovnávanie	19
3 Modelovanie airglow-u pomocou hlbokého učenia	20
3.1 Analýza súčasného stavu	20
3.2 Pochopenie cieľa	22
3.3 Pochopenie dát	24
3.4 Príprava dát	27
3.5 Explorácia algoritmu	28
3.6 Výsledné modely a vyhodnotenie	34
4 Záver	39
Zoznam príloh	44

Zoznam obrázkov

1-1	Pozorovanie airglow-u z porchu Mesiaca počas misie Apollo 16 a misiou GOLD. Zdroj: NASA	4
1-2	Snímky z misie GOLD pre Severnú pologuľu. Zdroj: NASA	6
1-3	Nákres snímača misie GOLD zobrazujúci hlavné časti snímača potrebné pre zachytenie údajov. Zdroj: NASA	8
1-4	Typická sekvencia nočného skenovania. Na začiatku noci je použitý kanál B na snímanie oboch pologúľ. Pri posledných 4 snímkoch bol použitý aj kanál A na snímanie južnej pologule zatiaľ, čo kanál B stále snímal severnú. Časové rozmedzia medzi snímkami sú 15 minút. Zdroj: NASA	11
2-1	Diagram GAN procesu pri učení sa generovania číslic. Zdroj: (Silva, 2020)	15
2-2	Architektúra U-net – príklad pre 32x32 pixelov v najnižšom rozlíšení. Zdroj: (Ronneberger et al., 2015)	17
3-1	Výsledky grafického zobrazenia dát. Prvý obrázok zobrazuje žiarenie 1356. Druhý zobrazuje žiarenie LBH. Na treťom obrázku je možné vidieť SZA. Posledný obrázok znázorňuje celé zobrazenie pre žiarenie 1356 bez podkladu Zeme.	26
3-2	Grafické zobrazenie prekrývania dát	27
3-3	Porovnanie predpovedaných (horná časť) a reálnych dát (dolná časť) našich modelov.	37
3-4	Porovnanie predpovedanej (horná časť) a reálnej (dolná časť) sekvencie severného modelu pre 20.1.2020	37

Zoznam tabuliek

3-1 Zoznam premenných v dátach z misie GOLD ktoré boli dôležité pri tvorbe grafického zobrazenia	25
3-2 Výsledky metrík 5 nasledujúcich snímok pri použití sekvencie 5 snímok pri vytváraní výsledkov modelu	31
3-3 Výsledky metrík 5 nasledujúcich snímok pri použití sekvencie 2 snímok pri vytváraní výsledkov modelu	32
3-4 Výsledky metrík 5 nasledujúcich snímok pri použití nastavenia -- <code>batchsize 5</code>	33
3-5 Výsledky metrík 5 nasledujúcich snímok pri použití nastavenia -- <code>fps 1</code>	33
3-6 Výsledky metrík 5 nasledujúcich snímok pri použití nastavenia -- <code>no_flip</code>	34
3-7 Výsledky metrík 6 po sebe idúcich snímok pre Severnú a Južnú pologuľu pri rôznych vybraných dňoch Januára 2020	36

Slovník skratiek

MSE Mean Squared Error

PSNR Peak Signal-to-Noise Ratio

NASA National Aeronautics and Space Administration

GOLD The Global-scale Observations of the Limb and Disk

UV Ultraviolet radiation

EUV extreme ultraviolet

GAN Generative Adversarial Networks

SPDF NASA's Space Physics Data Facility

Úvod

Hlboké učenie je v dnešnej dobe čoraz rozšírenejšia metóda analýzy rôznych dát, ktorú využívame každý deň a ani si to možno neuvedomujeme. Uplatnenie nachádza aj pri vedeckom výskume, a to najmä pri práci s veľkým množstvom dát, ktoré by nebolo inak možné spracovať alebo pri predikcii rôznych javov. Jednou oblasťou, kde sa predikovanie začína používať čoraz častejšie je oblasť atmosféry Zeme.

Vesmír je v dnešnej dobe stálou veľkou neznámou a je potrebné urobiť ešte veľké množstvo experimentov a skúmaní na to, aby sme pochopili mnohé jeho procesy a udalosti v ňom. Inak tomu nie je ani na jeho rozmedzí so Zemou. Horná atmosféra našej planéty je stále málo preskúmaná oblasť s množstvom rôznych úkazov, ktoré je potrebné skúmať.

Jedným z nich je aj žiarenie pomenované airglow. Toto žiarenie je možné pozorovať aj zo Zeme zvyčajne počas úplne tmavej noci. Zo Zeme je tento jav skúmaný pomocou rôznych observatórií no najlepšie je pozorovateľný pomocou satelitov. Takýmto skúmaním sa zaoberá aj NASA misia GOLD. Táto misia zbiera o airglow dáta každý deň počas dňa aj noci pomocou snímača na obežnej dráhe Zeme. Tieto dáta sú následne verejne dostupné no keďže je táto misia ešte relatívne mladá nebolo zatiaľ vydaných veľa článkov, ktoré by s týmito dátami pracovali.

Cieľom našej práce, teda bude spracovať dáta z tejto misie do grafickej podoby, nájsť vhodný algoritmus hlbokého učenia a následne ho aplikovať na tieto snímky, vďaka čomu preskúame či je možné pomocou hlbokého učenia rozšíriť pozorovania airglow-u, a tak detailnejšie sledovať jeho vývoj.

1 Doména skúmania

1.1 Airglow

Počas noci sa dá na oblohe pozorovať množstvo svetelných úkazov, ako napríklad hviezdy, Mesiac alebo planéty. Ak by sme všetky tieto svetelné objekty odstránili z nočnej oblohy stále by však obloha nebola kompletne čierna. Mohli by sme pozorovať jemné farebné žiarenie nachádzajúce sa v hornej atmosfére Zeme.

Horná atmosféra Zeme sa nachádza vo výškach 80 km od zemského povrchu a jej základnými časťami sú atómy a molekuly dusíka, kyslíka, vodíka a hélia. Táto oblasť reaguje na zmeny nielen zo Zeme spôsobené počasím, ale aj na zmeny prichádzajúce z vesmíru, najmä slnečným žiarením. Práve táto oblasť je vhodná pre vznik rôznych fotochemických úkazov pričom jedným z nich je airglow. Časť atmosféry, v ktorej sa toto žiarenie tvorí sa nazýva termosféra a ionosféra, ktorá žiari nie len vo viditeľnom svetle, ale aj v ultrafialovej a infračervenej časti spektra. Je veľmi ťažké ju pozorovať zo zemského povrchu, pretože sú potrebné podmienky, kde nie je takmer žiadne svetelné znečistenie, čiže je potrebná veľmi tmavá noc. Airglow je však dobre viditeľný pri pohľade na zem z vesmíru, kde ho môžeme pozorovať medzi povrchom zeme a začiatkom vesmíru. Toto žiarenie bolo pozorovateľné a zachytené v roku 1972 z povrchu Mesiaca pomocou UV kamery, čo je možné vidieť na Obrázku 1–1. Bolo zistené, že takéto žiarenie nie je výnimočné iba pre Zem, ale tento úkaz bol taktiež pozorovaný na iných planétach, ako napríklad Mars alebo Venuša. V dnešnej dobe sa týmto fenoménom na našej zemi zaberá napríklad NASA misia GOLD (z angl. *The Global-scale Observations of the Limb and Disk*) ktorá bude predstavená v časti 1.2 (Savigny, 2017).

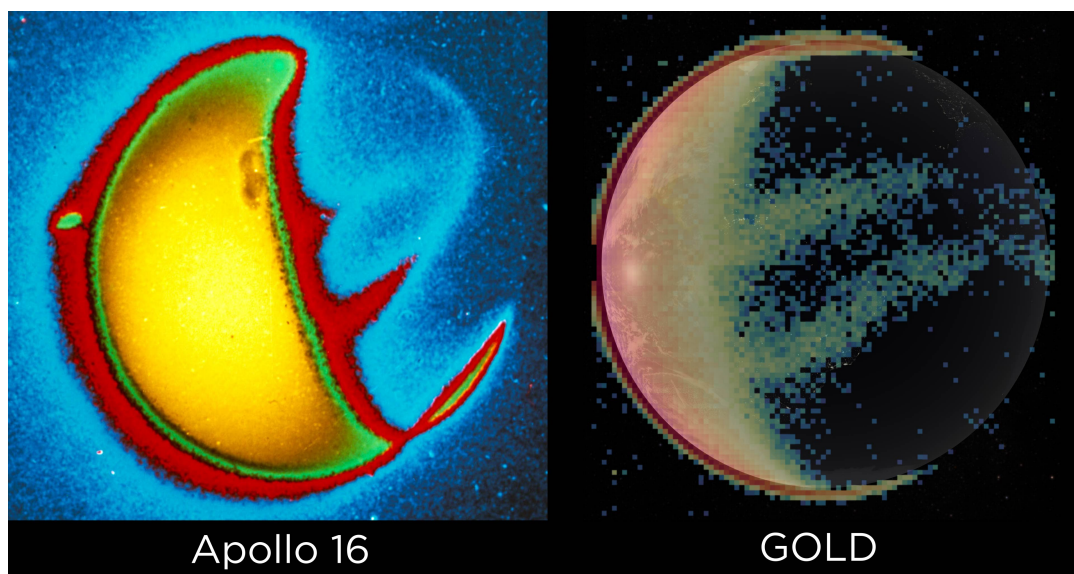
Airglow je spôsobovaný absorpciou slnečných ultrafialových a röntgenových lúčov molekulami vzduchu. Väčšina tejto žiary sa nachádza v oblasti od asi 80 až 300 km nad povrchom Zeme, pričom najjasnejšia oblasť je sústredená vo výškach okolo 97 km. Žiarenie je spôsobené chemickými reakciami slnečného žiarenia s atómami a molekulami prítomnými vo vyšších vrstvách atmosféry. Slnečné svetlo do-

dáva energiu potrebnú na excitáciu atómov a tie následne vytvárajú žiarenie na určitých vlnových dĺžkach. V tejto oblasti vedci najčastejšie pozorujú emisie sodíka (Na), hydroxylových radikálov (OH), molekulárneho kyslíka (O₂) a atómového kyslíka (O). Emisie sodíka sa vyskytujú vo výškach približne 80 až 100 km nad zemským povrchom, zatiaľ čo emisie OH, molekulárneho kyslíka a atómového kyslíka sa vyskytujú najčastejšie vo výškach 87 km až 95 km. Žiarenie týchto molekúl a atómov možno pozorovať vo viditeľnej časti elektromagnetického spektra. Pri emisiách sodíka ide o vlnovú dĺžku približne 590 nm, teda žiaria na žltoranžovo. Vlnové dĺžky emisií z OH a molekulárneho kyslíka však pokrývajú omnoho širšie pásma, a to približne od 650 do 1800 nm (červená a IR) a 120 až 490 nm (UV až modrá). Emisie atómového kyslíka sa vyskytujú v troch rôznych vlnových dĺžkach, ktoré sa nachádzajú pri 558 nm (zelená), 630 nm a 632 nm (červená) v rámci elektromagnetického spektra. Airglow je podobný polárnej žiare, avšak tá je spôsobená vysoko energetickými časticami pochádzajúcimi zo slnečného vetra zatiaľ čo airglow-u dodáva energiu bežné slnečné žiarenie. Pri porovnaní s polárnou žiarou ktorá sa vyskytuje len občasne a na pólach Zeme, airglow sa vyskytuje neustále v atmosfére vo forme vrstvy obklopujúcej celú našu planétu. Intenzita žiarenia je zhruba jedna desatina žiarenia hviezd, a teda je problémové ho pozorovať zo zemského povrchu bez použitia špecializovaných prístrojov (Tran, 2018).

Na Obrázku 1 – 1 je možné vidieť na obidvoch zobrazeniach farebné pásy smerujúce do nočnej časti našej Zeme. Tieto oblasti sa nachádzajú severne a Južne od rovníka a nazývajú sa Appletonova anomália. Ide o miesta, kde Zemské magnetické póly pôsobia na hornú atmosféru pričom spôsobujú husté fontány plazmy (Lühr and Rother, 2003) .

Airglow sa delí podľa závislosti uhlu dopadu slnečných lúčov na:

- **Dayglow** – nastáva, keď je pozorovaná plocha pod priamym dopadom slnečných lúčov, teda je osvetľovaná celá atmosféra našej planéty, pri pozorovaní zo zeme sa deje práve cez deň



Obrázok 1 – 1 Pozorovanie airglow-u z porchu Mesiaca počas misie Apollo 16 a misiou GOLD.
Zdroj: NASA

- **Twilightglow** – nastáva pri osvetlení len vrchnej časti atmosféry, pri pozorovaní zo zeme nastáva chvíľu po západe slnka a je taktiež najdlhšie pozorovateľná, pretože horná atmosféra je stále osvetľovaná zatiaľ, čo pozorovateľ je už v tme
- **Nightglow** – nastáva, keď je celá atmosféra v tme, a teda slnečné lúče sú pod uhlom menším ako 18° , teda nastáva takzvaná astronomická noc.

Pri pozorovaní z ISS alebo satelitu na obežnej dráhe Zeme je možné tieto delenia pozorovať súčasne keďže Slnko je pod iným uhlom pre rozdielne časti Zeme (Case, 2021).

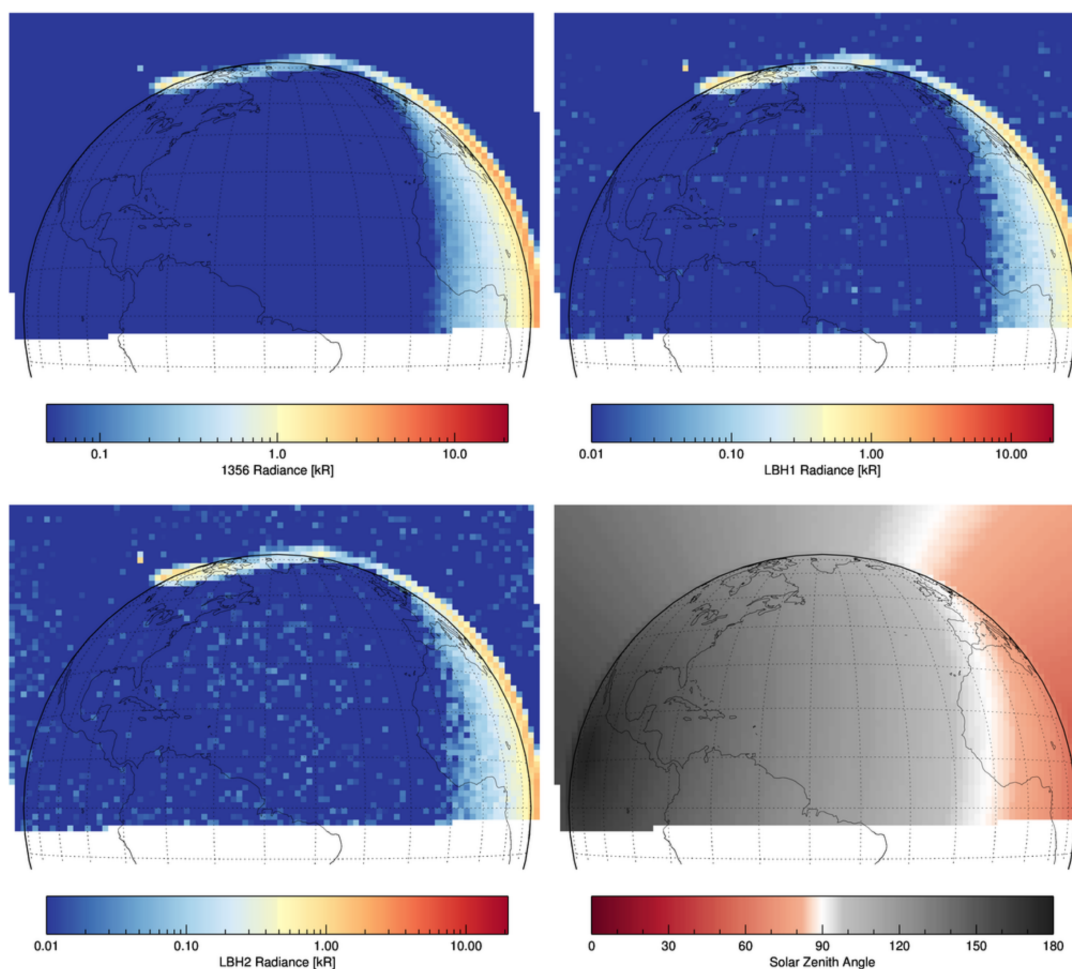
1.2 NASA misia GOLD

Pochopenie globálneho systému termosféry a ionosféry je veľmi dôležité pre lepšie pochopenie vzťahu medzi vesmírom a atmosférou našej Zeme. Tuto oblasť skúmalo niekoľko vesmírnych misií pričom boli zamerané hlavne na to, ako táto oblasť samotná reaguje na rôzne zmeny. Misia Globálneho pozorovania Zemského disku a jeho

okraja (z angl. *The Global-scale Observations of the Limb and Disk - GOLD*) má za úlohu určiť, ako reaguje počasie v tejto oblasti, čím pomáha k lepšiemu pochopeniu vzťahu medzi vesmírom a atmosférou Zeme. Misia bola spustená 25. Januára 2018 organizáciou NASA (Národný úrad pre letectvo a vesmír, skratka z angl. *National Aeronautics and Space Administration*). GOLD sa zameriava na zodpovedanie štyroch hlavných otázok o reakciách termosféry a ionosféry na sily pôsobiace zo strany Zeme ako aj Vesmíru. Pre odpoveď na tieto otázky používa misia ultrafialový (UV) snímač v vysokej kadencii, ktorý sa nachádza na satelite umiestnenom na geostacionárnej obežnej dráhe okolo Zeme vo výške približne 35400 km. Toto umiestnenie umožňuje časté a opakované snímanie, ktoré pokrýva takmer celú hemisféru nad Severnou a Južnou Amerikou na tých istých miestach počas celého dňa. Ukážku snímania nad Severnou pologoulou je možné vidieť na Obrázku 1–2 (Lewin, 2018).

Prvou z hlavných otázok je, ako geomagnetické búrky menia teplotu a štruktúru zloženia termosféry. Predošlé skúmania búrok v tejto oblasti mali limitované možnosti skúmania, a to hlavne z dôvodu, že im chýbala možnosť rozlišovať medzi časom v mieste pozorovania a zmeny zemepisnej dĺžky v globálnom meradle vďaka čomu bolo zložité porozumieť zmene v meranej oblasti. Merania z povrchu Zeme a satelitov v nízkej Zemskej orbite majú zasa príliš obmedzené možnosti pre rozšírenie nášho doterajšieho chápania. Misia GOLD po prvý krát priniesla rozsiahle zobrazenie zmien v teplote a zložení vyvolaných geomagnetickou búrkou. Tieto zobrazenia pomáhajú odpovedať na problémy napríklad, ako sa mení vertikálna štruktúra O₂ pri geomagnetických búrkach, či počiatková teplota a distribúcia molekúl ovplyvňuje reakciu na búrky a, či obnovenie pozorovanej oblasti je ovládané teplotou alebo zložením atmosféry. Vďaka skenovaniu rovnakej oblasti je taktiež možné porovnať zmeny počas búrok, ktoré sa objavujú približne 10 krát za rok s pozorovaniami mimo týchto búrok.

Ďalšou otázkou, ktorou sa zaoberá misia GOLD je celosvetová reakcia termosféry na slnečnú extrémnu ultrafialovú variabilitu. Slnečné žiarenie je jednou z hlavných príčin zmien v klíme termosféry. Ultrafialové žiarenie a röntgenové lúče ionizujú ne-



Obrázok 1 – 2 Snímky z misie GOLD pre Severnú pologuľu. Zdroj: NASA

utrálny plyn vďaka čomu sa vytvára nízka a stredná ionosféra. Hlavnou spektrálnou oblasťou pre skúmané miesta je od približne 1 nm do 103 nm (extrémne ultrafialové mäkké röntgenové žiarenie). Táto oblasť sa označuje ako EUV. Pomocou meraní misie GOLD je možné rozdeliť priestorové a časové účinky zo žiarení aj počas slnečných erupcií. Merania pomôžu objasniť problémy ako napríklad aká bude odozva termosférickej teploty a zloženia na pôsobenie slnečného EUV alebo či je termosférická odozva na erupcie úmerná meranej slnečnej emisii.

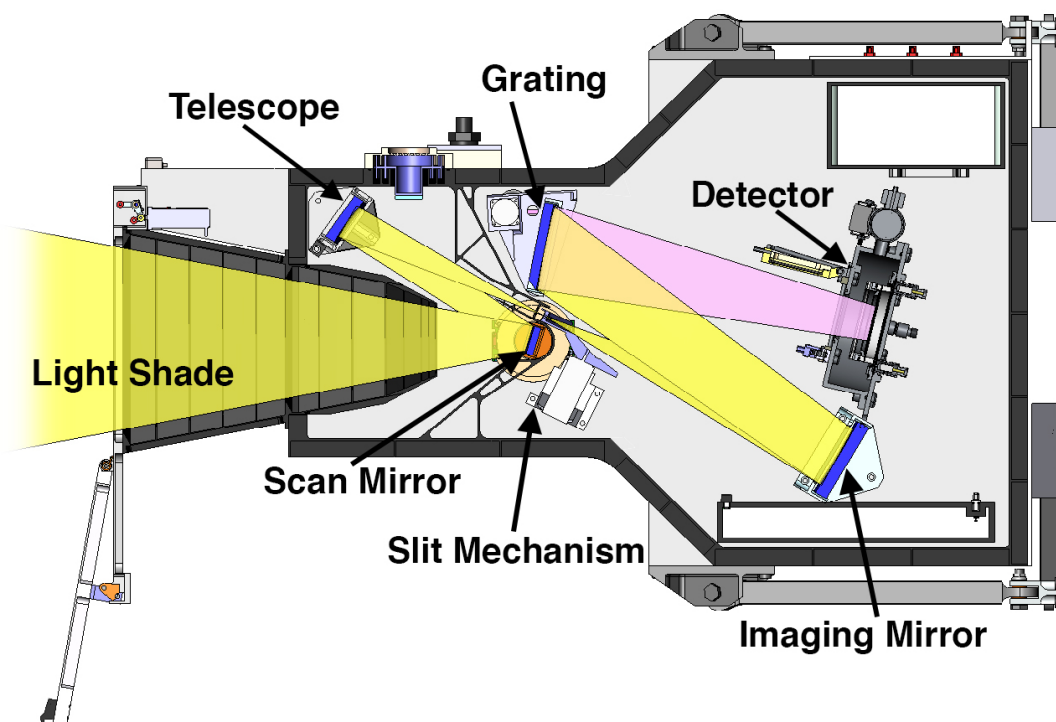
Tretou otázkou je aké významné sú účinky atmosférických vln a prúdení šíriacich sa zdola na štruktúru termosférickej teploty. Atmosférické veľkoškálové prúdenia sú vlny, ktorých periodicita zodpovedá 24 hodinám, ich rozsah je globálny a sú hlavným

zdrojom zmien vo vetre a teplotách v spodnej termosfére pričom sú veľmi dôležité pre vznik ionosférických elektrických polí. Na rozdiel od pozorovaní z nízkej orbity Zeme, misia GOLD je schopná pozorovať časové a priestorové zmeny súčasne, pričom pozoruje meranú oblasť na Zemi a zmeny premenlivosti prúdenia počas niekoľkých dní za sebou. Takéto rozšírené zbieranie dát z miesta merania je podobné systému, ktoré používajú pozemné pozorovania, avšak GOLD má možnosť pozorovať naraz omnoho väčšiu geografickú plochu, vďaka ktorej je možné získať rovnaké pozorovania pre pevninu aj oceán. Pri tejto otázke sa GOLD zaoberá hlavne tým aká je denná a sezónna premenlivosť vln v strednej termosfére.

Poslednou otázkou je, ako ovplyvňuje nočná rovníková ionosféra vznik a vývoj nepravidelností hustoty rovníkovej plazmy. Nepravidelnosti hustoty plazmy v nízkych zemepisných šírkach spôsobujú takzvané ionosférické scintilácie, ktoré sú tiež známe, ako „ionosférické bubliny“ a sú produkované v ionosfére po západe slnka Rayleigh-Taylor (R-T) nestabilitou. Miera objavovania týchto bublín je spätá so stavom ionosféry nízkej zemepisnej šírky, ktorá je ovplyvnená termosférickými prúdeniami, geomagnetickými búrkami a ďalšími procesmi, o ktorých zatiaľ veľa nevieme. Počas geomagnetických búrok, zmeny vo vysokej šírke elektrického potenciálu menia elektrodynamiku a rýchlosti presunov v oblasti rovníka, ktoré následne riadia výšku, intenzitu a separáciu rovníkových ionizačných anomálií. Atmosférické prúdenia a planetárne vlny tiež ovplyvňujú vertikálne presuny. GOLD meria vrcholovú elektrónovú hustotu v rovníkovej ionosfére v noci z pohľadu s pevnou zemepisnou dĺžkou vďaka čomu poskytuje reálnu časovú postupujúcu mapu na globálnej škále. Táto misia pri tejto otázke zodpovedá hlavne na to aké faktory v konfigurácii rovníkovej ionosféry ovplyvňujú tvorbu ionosférických bublín a, čo sa deje s ionosférickými bublinami, keď sa vyvíjajú a presúvajú (Eastes et al., 2017).

Kvôli požadovanej perspektíve vedeckých cieľov bolo potrebné umiestniť misiu GOLD na geostacionárnu orbitu okolo zeme. Misia GOLD nie je umiestnená na čisto výskumnom satelite, ale ako súčasť komerčného komunikačného satelitu SES-14 v spolupráci so spoločnosťou SES Government Solutions (SES GS) (*SES-14 Integrates*

NASA Ultraviolet Space Spectrograph, 2017).



Obrázok 1–3 Nákres snímača misie GOLD zobrazujúci hlavné časti snímača potrebné pre zachytenie údajov. Zdroj: NASA

Snímač misie GOLD bol skonštruovaný v Laboratóriu pre atmosférickú a vesmírnu fyziku v Colorade (*Quick facts: Global-scale observations of the limb and disk (gold)*, 2019) a obsahuje dve identické a navzájom nezávislé optické kanály, pričom oba sú schopné vykonávať tie isté merania. Každý kanál obsahuje ultrafialový spektrogaf vybavený zobrazovacím detektorom, ktorý pokrýva 132 až 162 nm. Táto vlnová dĺžka pokrýva hlavne emisie, a to atómový kyslík (135,6 nm) a molekulárny dusík (systém Lyman-Birge-Hopfieldovho – LBH pásma 132 – 162 nm). Hlavné optické komponenty snímača sú zobrazené na Obrázku 1–4. Súčasťou prístroja je taktiež jedno-zrkadlový ďalekohľad. Počas pozorovania, presný mechanizmus otáča zrkadlo ďalekohľadu tak, aby sa obraz pohyboval od východu na západ cez pozorovanú zemskú atmosféru. Štrbina však nie je dostatočne vysoká pre pokrytie celej plochy Zeme, ale dokáže snímať iba určitú časť jednej pologule (buď severnú alebo

južnú). To sa dosiahne naklonením zrkadla v ďalekohľade o približne $4,5^\circ$. Pri takomto nastavení jedna strana zrkadla premieta časť severnej pologule a druhá strana časť južnej pologule. Hlavná životnosť tejto misie, a teda aj snímača bola očakávaná na dva roky pričom oba kanály doteraz vykonávajú merania potrebné na splnenie základných vedeckých otázok tejto misie a poskytujú údaje dostatočné na určenie sezónnej premenlivosti v termosfére a ionosfére (*Gold Instrument*, n.d.).

1.3 Pozorovania misie

Zo snímok misie GOLD sa odvodzujú simultánne snímky termosférickej teploty a pomerov neutrálneho zloženia s polhodinovou kadenciou počas denných pozorovaní na celom disku. V noci GOLD zobrazuje ionosféru nízkej šírky, kde sú často pozorované oblasti so zníženou hustotou spôsobenou nestabilitou alebo nepravidelnosťami, ktoré sa tvoria po súmraku. Sú 3 hlavné oblasti zemskej atmosféry, o ktorých misia GOLD zbiera informácie: osvetlená časť priamym slnečným žiarením, nočná atmosféra a okraj zemského disku. Pri týchto miestach sa získavajú tieto špecifické merania:

- **Denný disk**

- Snímky emisií kyslíka 135,6 nm a N2 LBH nad $\pm 60^\circ$ zemepisnej šírky a $\pm 70^\circ$ zemepisnej dĺžky v porovnaní s nadirom satelitu
- Snímky s termosférickou teplotou pri 60-minútovej kadencii, priestorovým rozlíšením 250 km na 250 km (v najnižšom bode) a presnosťou ± 55 K
- Snímok pomeru hustoty stĺpca O/N2 pri 30-minútovej kadencii, priestorové rozlíšenie 250 km na 250 km (pri najnižšej hodnote) a presnosť s hodnotou 10 tohto pomeru

- **Nočný disk**

- Snímky Nmax F2 na vrchole rovníkových oblúkov s presnosťou 10 % a rozlíšením zemepisnej šírky 2°

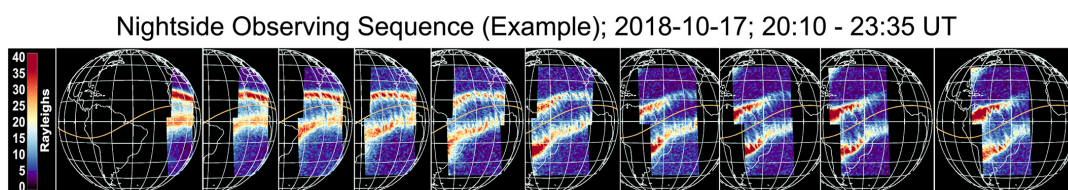
- Sledovanie ionosferických bublín v rámci rovníkového oblúka s 100-kilometrovým priestorovým rozlíšením (v spodnej časti) v pozdĺžnom smere

- **Okraj disku**

- Meranie exosférickej teploty v blízkosti rovníka z emisií N₂ LBH s presnosťou ± 40 K
- Meranie hustoty stĺpcov molekulárneho kyslíka (O₂) pri vertikálnom rozlíšení 10 km a s presnosťou ± 10 % nad 150 km od hviezdnych zákrytov

Mimo oblastí polárnej žiari je najvýraznejšie pozorovateľná airglow emisia kyslíka 135,6 nm a je pozorovateľná hlavne v oblasti rovníka počas noci. Hustota ionosfery v tejto oblasti, od ktorej závisí jasnosť žiarenia, je zaujímavá pre výskum hlavne pre jej premenlivosť a štruktúru. Žiarenie vzniká radiačnou re-kombináciou atómových kyslíkových iónov, ktoré vznikajú spolu s elektrónmi na dennej strane pred západom slnka. Východné elektrické pole z ionosferického dynama ich ťahá cez siločiaru magnetického poľa do vyšších nadmorských výšok. V tejto oblasti je strata atómových iónov kyslíka výmenou náboja nevýznamná a dominujú tu výrazne pomalšie procesy atómovej straty. Gravitačné sily tiež ovplyvňujú tok a pridávajú sa k súčasnému toku iónov smerom k pólu. Počas skenovania nočnej oblohy sa používa režim čiastočného skenovania disku tak, aby mali výsledky lepší pomer signálu k šumu pri porovnaní so scénami celého disku. Oproti skenovaniu celého disku je zmenený v 3 hlavných hladiskách. Prvá je, že sa pri skenovaní používa širšia štrbina s nižším rozlíšením, čím sa dosiahne dvojnásobná reakcia meracích prístrojov na emisiu airglow-u. Druhá spočíva v predĺžení pozorovacieho času určitého miesta vďaka skenovaniu pri približne 45° zemepisnej dĺžky (pre väčšinu nočných skenovaní blízko západu slnka). Kadencia skenovania je oproti ostatným skenovaniam nezmenená, a to každých 15 minút. Treťou zmenou oproti snímaniu celého disku je, že nočné skenovanie nie je prerušované skenovaním iných oblastí. Počas noci začína skenovanie pomocou kanála B vždy o 20:10 UT, striedavo medzi Severnou a južnou pologouľou. Od 23:10 do

00:40 UT sa pridá aj kanál A ktorý skenuje severnú pologuľu zatiaľ, čo kanál B skenuje južnú pologuľu. Všetky dáta z nočného skenovania sú zoskupené do konštantnej uhlovej mriežky s priestorovým rozlíšením približne 93 km v najnižšom bode. Nočné dáta sú označené ako NI1. Príklad dostupnosti dát z nočného skenovania môže byť viditeľný na Obrázku 1–4 (Eastes et al., 2020).



Obrázok 1–4 Typická sekvencia nočného skenovania. Na začiatku noci je použitý kanál B na snímanie oboch pologúľ. Pri posledných 4 snímkoch bol použitý aj kanál A na snímanie južnej pologule zatiaľ, čo kanál B stále snímal severnú. Časové rozmedzia medzi snímkami sú 15 minút. Zdroj: NASA

2 Hlboké učenie

2.1 Strojové učenie

Strojové učenie je metóda analýzy dát, pri ktorej sa automatizuje vytváranie analytických modelov. Ide o odvetvie umelej inteligencie založené na predpoklade, že systémy sa dokážu učiť z vstupných údajov, pričom v nich dokážu identifikovať vzory a taktiež robiť rozhodnutia s minimálnym zásahom človeka. Analyzuje a skúma konštrukciu algoritmov, ktoré sa dokážu učiť a predpovedať podľa vstupných dát. Strojové učenie dokáže riešiť problémy rýchlosťou a rozsahom, omnoho lepším ako akýkoľvek človek. Vďaka obrovskému množstvu výpočtových schopností je možné stroje trénovať na identifikáciu vzorcov a vzťahov medzi vstupnými údajmi.

Strojové učenie, ako koncept existuje už pomerne dlhú dobu. Prvá zmienka, a teda aj zavedenie termínu strojové učenie sa pripisuje Arthurovi Samuelovi, ktorý bol počítačový vedec z IBM a priekopník v oblasti AI a počítačových hier. Samuel navrhol počítačový program, v ktorom bolo možné hrať stolovú hru dáma. Čím dlhšie tento program dámu hral, tým viac sa učil z toho, čo sa udialo a taktiež aj zo svojich chýb, ktoré urobil. Tieto skúsenosti následne použitím algoritmov využíval na vytváranie predpovedí, ako by mal pri hraní dámy pokračovať (Samuel, 2000).

Definovanie pojmu strojové učenie však nebol úplný začiatok rozmýšľania o samo mysliacich počítačoch/strojoch. Desiatročia pred touto definíciou bola uverejnená práca od Alana Turinga, s názvom Výpočtové stroje a inteligencia (Computing Machinery and Intelligence). (Turing, 1950) Túto prácu publikoval v roku 1950 a je považovaná za zrodenie konverzácie o umelej inteligencii. Turing, často označovaný aj ako „otec informatiky“, si v tejto práci kladie otázku: „Dokážu stroje myslieť?“. Z tohto článku taktiež vznikol test, ktorý je dodnes známy ako Turingov test. Ide o test, kde sa osoba pokúša rozlíšiť medzi textovou odpoveďou počítača a ľudskou odpoveďou. Aj, keď tento test prešiel od svojho zverejnenia mnohými modifikáciami, zostáva dôležitou súčasťou histórie umelej inteligencie.

Strojové učenie sa najčastejšie delí podľa spôsobu učenia na kontrolované (angl.

supervised learning), nekontrolované (angl. *unsupervised learning*) a učenie odmenou a trestom (angl. *reinforcement learning*):

- **Kontrolované učenie**

Algoritmy kontrolovaného učenia využívajú na učenie a analýzu vopred označené dáta, ktoré boli označené buď analýzou iného algoritmu alebo boli označené človekom. Analýzou vopred označených údajov algoritmus vytvára odvodenú funkciu na predpovedanie výstupných hodnôt. Pri neurónových sieťach sa v praxi využíva algoritmus spätného šírenia chýb (angl. *backpropagation*) (HECHT-NIELSEN, 1992). Ako z názvu vyplýva vďaka tomuto algoritmu sa v procese učenia spätne aktualizujú váhy, čo zabezpečuje proces učenia sa. Tento spôsob učenia má taktiež výhodu, že nad ním máme väčšiu kontrolu keďže vieme skutočnú hodnotu alebo triedu predikcií (Singh et al., 2016).

- **Nekontrolované učenie**

Algoritmy nekontrolovaného učenia sa používajú, keď dostupné dáta, ktoré budú použité na tréningovanie nie sú predom označené alebo klasifikované. Algoritmus tohto učenia sa snaží zistiť, ako môžu systémy odvodiť funkciu na opis skrytej štruktúry z neoznačených údajov, teda sa snaží nájsť akékoľvek súvislosti v dátach. Systém v žiadnom bode nepozná s istotou, ako má vyzerat správny výstup. Namiesto toho vytvára predikcie z množín údajov o tom, aký by mal byť výstup. Výhody tohto spôsobu sú jeho rýchlosť a taktiež možnosť použitia na dáta, ktoré predtým neboli označované (Hahne et al., 2008).

- **Učenie odmenou a trestom**

Algoritmy učenia odmenou a trestom sú metódy, ktoré interagujú s prostredím vytváraním akcií a objavovaním chýb alebo odmien. Táto metóda umožňuje algoritmom automaticky určiť ideálne správanie v konkrétnom kontexte, aby sa maximalizoval jeho výkon. Jednoduchá spätná väzba odmeny je potrebná na to, aby algoritmus zistil, ktorá akcia je najlepšia (Kaelbling et al., 1996).

Hlboké učenie je odvetvie strojového učenia, ktoré využíva neurónové siete s veľkým počtom vrstiev. Hlboká neurónová sieť analyzuje údaje podobne, ako by sa na problém pozerali ľudia. V tradičnom strojovom učení má algoritmus k dispozícii súbor relevantných funkcií na analýzu. Pri hlbokom učení však algoritmus dostane nespracované údaje a sám sa rozhodne, ktoré funkcie a súvislosti sú relevantné. Siete hlbokého učenia často dosahujú lepšie výsledky pri zvýšení množstva dát určených na tréning. Hlboké učenie sa v podstate snaží napodobňovať fungovanie ľudského mozgu. Rovnako, ako sa ľudia učia zo skúseností aj algoritmus hlbokého učenia môže vykonávať úlohu opakovane, pričom sa zlepšuje v presnosti výsledkov. Hlboké učenie je možné použiť pri širokom množstve problémov, pri ktorých bolo potrebné akékoľvek skúmanie (LeCun et al., 2015).

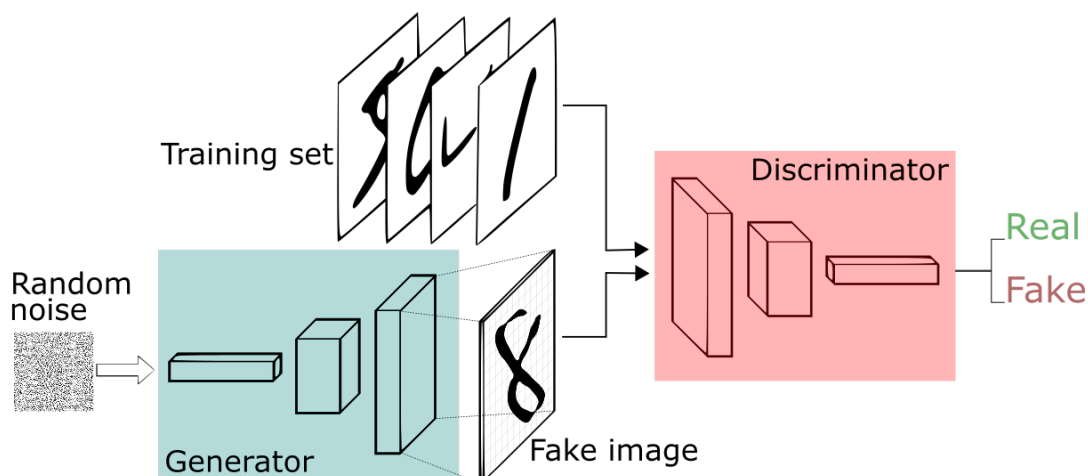
2.2 Generatívne adverzné siete

Generatívne adverzné siete (označované skrátene ako GAN z anglického *Generative Adversarial Networks*) sú prístupom ku generatívne modelovaniu pomocou metód hlbokého učenia, ako sú napríklad konvolučné neurónové siete. Generatívne modelovanie je typ strojového učenia, ktoré používa učenie sa bez dohľadu a zahŕňa automatické objavovanie a učenie sa zákonitostí alebo vzorcov vo vstupných údajoch. Výsledný model možno použiť na generovanie alebo výstup nových príkladov, ktoré sa s určitou presnosťou podobajú na pôvodné vstupné údaje. GAN používajú zaujímavý spôsob tréningu generatívneho modelu, pretože pracujú kvázi na princípe učenia s dohľadom, aj keď takéto dáta nedostanú. Toto je docielené vďaka dvom podmodelom. Jedna neurónová sieť, nazývaná generátor, generuje nové dátové inštancie, zatiaľ čo druhá, diskriminátor, vyhodnocuje ich pravosť.

Diskriminátor, teda rozhodne či každá inštancia údajov, ktoré kontroluje, patrí do pôvodného súboru údajov o tréningu alebo nie. Jednoduchým príkladom takéhoto generovania môže byť napríklad generovanie ručne písaných číslíc, pričom vstupné dáta sú reálne čísllice písané rukou (príkladom takéhoto súboru údajov môže byť

napríklad dataset MNIST (Deng, 2012)). Cieľom diskriminátora, keď je zobrazený príklad zo skutočného súboru údajov, je rozpoznať tie, ktoré sú autentické. Generátor medzitým vytvára nové, syntetické obrázky, ktoré odovzdáva diskriminátoru. Robí tak v nádeji, že aj oni budú považovaní za autentické, hoci sú falošné. Grafické znázornenie tohto procesu je možné vidieť na Obrázku 2–1.

GAN sú rýchlo sa meniacou oblasťou, ktorá sa pokúša generovať realistické príklady naprieč celým radom problémových domén, najmä v úlohách prekladu z obrázka na obrázok, ako je preklad fotografií z leta na zimu alebo z dňa na noc a pri vytváraní fotorealistických fotografií predmetov, scén a ľudí, o ktorých ani ľudia nevedia rozoznať, že sú falošné (Creswell et al., 2018).



Obrázok 2–1 Diagram GAN procesu pri učení sa generovania čísiel. Zdroj: (Silva, 2020)

2.2.1 Pix2Pix model

Pix2Pix je GAN model určený na všeobecné účely prekladu obrázka na obrázok. Pix2Pix GAN je implementáciou podmienených GAN známich aj ako cGAN (Mirza and Osindero, 2014), ktoré sú rozšírením architektúry GAN, ktorá poskytuje kontrolu nad vygenerovaným obrázkom, napr. umožňuje vygenerovať obraz danej triedy. Ako taký musí byť Pix2Pix GAN nie len tréňovaný na súbore vstupných obrázkov, ale pre vytvorenie predikcie sú potrebné aj výstupné, teda cieľové obrázky, na ktoré

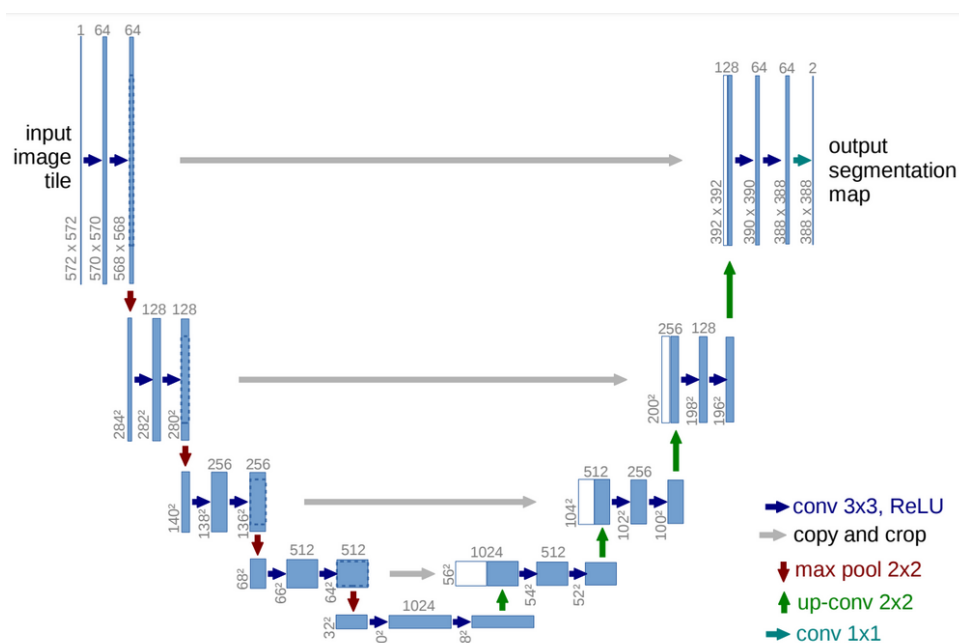
bude nadviazaná výsledná predikcia. Táto všeobecná architektúra umožňuje model Pix2Pix trénovať na celý rad úloh prekladu z obrázka do obrázka a taktiež použitie pri predpovedi nasledujúcich obrázkov sekvencie. Architektúra Pix2Pix sa skladá z 2 hlavných častí:

- **Generátor**

Ako generátor používa Pix2Pix model U-net. Architektúra U-net je symetrická v tvare písmena U a pozostáva z dvoch hlavných častí. Ľavá časť tvorí všeobecný konvolučný proces zatiaľ, čo pravá časť je expanzívna dráha, ktorú tvoria transponované 2D konvolučné vrstvy. Obraz je vložený do siete naľavo hore na Obrázku 2–2. Dáta sa následne šíria všetkými možnými cestami a výsledkom je mapa segmentácie. Hlavnou myšlienkou je doplniť bežnú kontrakčnú sieť o postupné vrstvy, kde operátori prevzorkovania nahrádzajú operácie združovania. U-Net sa učí segmentáciu v end-to-end nastavení. Pri vložení nespracovaného obrázku ako vstup, na konci získame mapu segmentácie. Vykonáva klasifikáciu každého pixelu, takže vstup a výstup má rovnakú veľkosť (Ronneberger et al., 2015).

- **Diskriminátor**

Na rozdiel od tradičného modelu GAN, ktorý používa hlbokú konvolučnú neurónovú sieť na klasifikáciu obrázkov, model Pix2Pix používa PatchGAN. (Demir and Ünal, 2018) Ide o hlbokú konvolučnú neurónovú sieť navrhnutú tak, aby klasifikovala časti vstupného obrazu ako skutočné alebo falošné, narušenie od vyhodnocovania celého obrázku. Diskriminačný model PatchGAN je implementovaný ako hlboká konvolučná neurónová sieť, ale počet vrstiev je nakonfigurovaný tak, aby sa efektívne prijímacie pole každého výstupu siete mapovalo na špecifickú veľkosť vo vstupnom obraze. Výstupom siete je jedna funkčná mapa skutočných a falošných predpovedí, z ktorých možno spriemerovať jedno skóre (Wang et al., 2017).



Obrázok 2–2 Architektúra U-net – príklad pre 32x32 pixelov v najnižšom rozlíšení. Zdroj: (Ronneberger et al., 2015)

2.3 Použité knižnice

Pre prácu s dátami sme zvolili programovací jazyk Python. Tento jazyk zaznamenal v poslednom desaťročí obrovský nárast popularity pre využitie pri strojovom učení vďaka veľkému množstvu dostupných knižníc strojového učenia a hlbokého učenia. Jedným z hlavných zameraní Pythonu je čitateľnosť a je taktiež všeobecne uznávaný za to, že sa ľahko učí, no v prípade potreby stále dokáže využiť silu programovacích jazykov na systémovej úrovni. Okrem výhod samotného jazyka, komunita okolo dostupných nástrojov a knižníc robí Python obzvlášť atraktívnym pre pracovné zaťaženie v oblasti skúmania dát, strojového učenia a vedeckých výpočtov (Raschka, 2015).

Pri našej práci s dátami sme používali nasledujúce základné knižnice:

- **netCDF4**

Knižnica netCDF4 umožňuje v Pythone jednoducho kontrolovať, čítať, agregovať a zapisovať súbory netCDF. NetCDF (z angl. *Network Common Data*

Form) je jedným z najdôležitejších formátov na ukladanie a zdieľanie vedeckých údajov. NetCDF4 ukladá údaje načítané z netCDF súboru do numPy poľa vďaka čomu ponúka prístup k celej syntaxi a funkciám numPy polí. Umožňuje taktiež extrahovať časť údajov zadaním indexov (Butler, 2018).

- **Matplotlib**

Knižnica Matplotlib je jednou z najvýkonnejších knižníc v pythone na vizualizáciu údajov. V tejto knižnici je možné vytvárať rôzne typy vizualizácií, ako sú čiarové grafy, bodové grafy, histogramy, stĺpcové grafy, koláčové grafy, krabicové grafy a mnoho ďalších rôznych grafov. Táto knižnica podporuje aj 3-rozmerné vykresľovanie (Hunter, 2007).

- **Basemap**

Basemap je nástroj založený na knižnici Matplotlib. Jeho úlohou je vykresľovanie 2D údajov na mapy v Pythone. Základná mapa sama o sebe nevytvára žiadne vykresľovanie, ale poskytuje priestor na transformáciu súradníc na jednu z 25 rôznych mapových projekcií. Matplotlib sa potom používa na vykresľovanie obrysov, obrázkov, vektorov, čiar alebo bodov v transformovaných súradniciach. Na mape sa nastavujú údaje o zobrazení pobreží, riek a hraníc spolu s metódami ich vykresľovania (Whitaker, 2011).

- **NumPy**

Knižnica NumPy (z anglického *numerical python* — číselný python) je široko používaná na prácu s číselnými poliami. NumPy dokáže spracovať viacrozmerné polia a matice spolu s veľkou zbierkou matematických operácií na prácu s týmito poliami. Pole zaberá menej pamäte a je jednoduchšie na použitie v porovnaní so zoznamami v Pythone. Okrem toho má taktiež vstavaný mechanizmus na špecifikáciu typov údajov. NumPy môže pracovať s jednotlivými prvkami v poli bez použitia slučiek a porozumení zoznamu (Harris et al., 2020).

2.4 Metriky pre porovnávanie

To či sa predpovedaný obraz rovná alebo podobá na pôvodný obraz môže byť dost subjektívne. Z tohto dôvodu je potrebné stanoviť kvantitatívne/empirické opatrenia na zistenie, ako veľmi sa očakávaný výsledný obraz podobá s tým ktorý vyprodukoval náš algoritmus. Pre tieto porovnania sú vhodné nasledujúce metriky:

- **MSE (z angl. *Mean Squared Error*)** – najbežnejšia metrika merania kvality obrazu. Meria množstvo chýb pri výsledkoch trénovaného modelu. Hodnotí priemernú druhú mocninu rozdielu medzi pozorovanými a predpokladanými hodnotami. V prípade, že pozorované obrázky nemajú žiadnu chybu (sú totožné), MSE sa rovná nule (Wang and Bovik, 2009). Vzorec pre MSE medzi dvoma obrázkami je nasledovný:

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (2.1)$$

kde x a y sú dva obrázky, pri ktorých $i = 1, 2, \dots, N$, kde N je počet vzoriek signálu (pixelov) a x_i a y_i sú hodnoty i -tých vzoriek v x a y , v tomto poradí.

- **PSNR (Peak Signal-to-Noise Ratio)** – pomer medzi maximálnou možnou hodnotou signálu a silou skresľujúceho šumu, ktorý ovplyvňuje kvalitu jeho zobrazenia. Pretože mnohé signály majú veľmi široký dynamický rozsah (pomer medzi najväčšou a najmenšou možnou hodnotou meniteľnej veličiny), PSNR sa zvyčajne vyjadruje v stupnici dekadického logaritmu. Pri tejto metrike je predmetom skúmania pomer medzi maximálnou hodnotou signálu a šumom (Wang et al., 2004). Vzorec pre PSNR je nasledovný:

$$PSNR = 20 \log_{10} \left(\frac{MAX_f}{\sqrt{MSE}} \right), \quad (2.2)$$

kde MAX_f je maximálna hodnota signálu, ktorá existuje v našom pôvodnom obrázku, o ktorom sa vie, že je dobrý.

3 Modelovanie airglow-u pomocou hlbokého učenia

3.1 Analýza súčasného stavu

Hlboké učenie je v dnešnej dobe použité v mnohých sférach nášho života. Či ide o spam filtre v našich emailových adresách alebo odomykanie našich telefónov pomocou našej tváre. Hlboké učenie má však veľký význam pre vedecké účely, kde je hlavne používané pri práci s veľkým množstvom dát, ktoré by inak nebolo možné spracovať alebo pri predpovedaní javov, pri ktorých by si človek súvislosť možno ani nevšimol. Jeden zo zaujímavých aspektov použitia strojového alebo hlbokého učenia je výskum vesmíru a našej atmosféry. V dnešnej dobe mnohé spoločnosti, ako napríklad NASA a Google, implementujú rôzne spôsoby hlbokého učenia na hľadanie nových nebeských telies, života na iných planétach a dokonca ich využívajú na zefektívnenie práce astronautov, keď sú vo vesmíre.

Ďalším príkladom využitia hlbokého učenia pre vesmírny výskum je spracovávanie dát zo satelitov okolo Zeme. Množstvo údajov produkovaných satelitmi krúžiacich okolo Zeme je obrovský. Technológie strojového učenia sa v dnešnej dobe už používajú na rozpoznávanie vecí, ako sú napríklad čierne diery alebo rádiové galaxie. Predtým, ako sa na štúdium údajov použili techniky strojového učenia, bolo potrebné toto množstvo dát spracovávať ručne, čo zaberalo príliš dlhú dobu. Pri skúmaní vesmíru, ale aj Zemskeho povrchu sú čoraz častejšie aplikované rôzne techniky strojového učenia. Tieto techniky majú veľký potenciál pomôcť vedcom riešiť niektoré z najnaliehavejších výziev a otázok, ktorým dnes čelíme, ako aj urýchliť celý proces skúmania a eliminovať ľudskú chybu, čo najviac (Russo and Lax, 2022).

Využitie metód hlbokého učenia pri skúmaní fenoménu airglow je v dnešnej dobe nie veľmi zaužívané a existuje len veľmi málo štúdií využívajúcich metódy hlbokého alebo strojového učenia pri skúmaní airglow-u. Tento fakt je taktiež dôvodom pre dôležitosť tejto práce, pretože vytvára prínos do oblasti, ktorá nie je dostatočne

preskúmaná. Pre bližšiu analýzu sme vybrali nasledujúce štúdie:

- **Automatická extrakcia tiažových vln z All-Sky airglow obrazu na základe strojového učenia** (Lai et al., 2019)

Táto práca sa zoberá spracovaním veľkého množstva dát pozemnej technológie airglow imager (ASAI) na štúdium atmosférických tiažových vln. Pre túto prácu vyvinuli program, ktorý ma za úlohu automaticky extrahovať vzory tiažových vln z obrázkov ASAI. Súčasťou extrakcie je taktiež klasifikačný model založený na konvolučnej neurónovej sieti (CNN) a model detekcie objektov. Ide o prvú štúdiu zaoberajúcou sa automatickou extrakciou vzorov tiažových vln v obrázkoch airglow. Program automatickej extrakcie, ktorý v tejto práci vyvinuli ponúka nový prístup k dlhodobej analýze charakteristík tiažových vln s vyššou účinnosťou a jednotnými štandardmi, aby sa predišlo náhodnej chybe v dôsledku manuálnych kontrol.

- **Modelovanie variability airglow-u počas obdobia troch slnečných cyklov** (Mackovjak et al., 2021)

Pri tejto práci išlo o preskúmanie či model strojového učenia možno úspešne použiť na modelovanie intenzity airglow-u na dátach, ktoré boli zachytené zo Zemskeho povrchu. Regresná úloha v tejto práci bola vykonaná pomocou časových radov kozmického počasia a parametrov ionosféry. Pri tejto práci boli použité techniky lineárnej regresie, neurónovej siete, random forest a XGBoost. Vyvinutý model založený na údajoch mal dobrú zhodu s bežne používaným modelom pre prácu s airglow s názvom GLOW a taktiež zachytával variácie airglow-u spôsobené cyklami slnečnej aktivity a zmenami ročných období. Výsledné modelované údaje tejto práce majú prispieť k pochopeniu procesov v oblasti rozhrania medzi vesmírnym prostredím a zemskou atmosférou.

Obe popísané práce pracujú síce s airglow dátami, a to konkrétne so sekvenciou dát, teda časovým radom. V prípade našej práce pracujeme so snímkami zo satelitov pozorujúcich airglow v Zemskej atmosfére, teda s obrazovými dátami.

Model Pix2PixHD, ktorý je použitý v tejto práci nemal zatiaľ použitie v oblasti skúmania vesmíru, teda týmto prinášame nový poznatok pri skúmaní airglow-u. Keďže ide o model pracujúci s predpovedaním videa je vhodný na použitie predpovedania časovo nadväzujúcich snímok zo satelitu. Mimo skúmania vesmíru bol model Pix2PixHD použitý napríklad pri práci s názvom Samo-kontrolovaná tanečná video syntéza podmienená hudbou (Ren et al., 2020). Cieľom práce bola automatická generácia realistického tanečného videa, ktoré zodpovedá rytmom hudby. Tento cieľ dosahujú vygenerovaním sekvencie ľudskej kostry z hudby a potom používajú naučené mapovanie polohy a vzhľadu na vytvorenie konečného videa. Čiže ide o vytváranie sekvencie snímok pomocou algoritmu Pix2PixHD.

3.2 Pochopenie cieľa

Cieľom tejto práce je pomocou hlbokého učenia vytvoriť model pre predpovedanie airglow-u v zemskej atmosfére. Tento model mal, ako vstupné dáta používať grafické zobrazenie snímaných hodnôt zo satelitu. Použité dáta pochádzajú z NASA misie GOLD, ktorá je relatívne nová misia keďže bola vypustená 25. Januára 2018 a prvé dáta boli odoslané na Zem 5. októbra toho istého roku. Aj napriek tomu, že dáta sú verejne dostupné už nejakú dobu nebolo zatiaľ vydané veľké množstvo prác, ktoré by do hĺbky skúmali tieto dáta a ešte menej ich pri tom používa hlboké učenie. Tento fakt je z časti spôsobený aj tým, že dáta boli v roku 2021 spätne upravené, aby obsahovali väčšie množstvo informácií. Z týchto dôvodov práca s dátami tejto misie bola náročnejšia keďže neexistujú rozsiahlejšie návody, ako s týmito dátami pracovať. Táto situácia je však taktiež príležitosťou vytvoriť výsledky, ktoré by mohli v budúcnosti pomôcť pri ďalšom skúmaní týchto dát. Oblasť vzniku airglow-u je taktiež v dnešnej dobe stále nie úplne preskúmaná, a teda by táto práca mohla pomôcť pri ďalšom výskume tejto oblasti. Pre dosiahnutie cieľa bolo potrebné urobiť nasledujúce kroky:

- **Stiahnuť dostatočné množstvo dát** – pre vytváranie modelov sme potre-

bovali získať väčšie množstvo dát, ktoré sú pri misii GOLD dostupné, a preto bude potrebné prísť na vhodný spôsob ich stiahnutia.

- **Pochopiť, ako sú dáta štruktúrované a, ako ich najlepšie zobrazíť na grafe** – s týmto typom údajov sme nemali žiadne predošlé skúsenosti, a tak bolo potrebné zistiť, ako vhodne pracovať s ich štruktúrou a, ako ich následne vykresliť tak, aby sa, čo najviac podobali zobrazeniu, ktoré je možné vidieť na stránke misie GOLD. Pri skúmaní práce s týmito dátami sme taktiež narazili len na jeden návod, ako s nimi pracovať, ktorý však nebol kompletný, a preto tento proces zabral najdlhšiu dobu z celej tvorby práce.
- **Vybrať vhodný spôsob predikcie** – keďže ide o prácu s grafickými dátami bolo potrebné vybrať aj odpovedajúci model, ktorý by dokázal s dátami pracovať no zároveň by nebol príliš zložitý na prácu s ním.
- **Vytvorenie grafických verzií pre každý dátový súbor vhodných pre vybraný spôsob strojového učenia** – cieľom tejto práce bola práca so snímkami zo satelitu, avšak dáta misie GOLD nie sú dostupné vo vhodnej grafickej podobe pre našu prácu, a tak sme si museli naša grafické dáta pripraviť sami.
- **Roztriedenie údajov podľa severnej a južnej pologule** – údaje sú snímané striedavo na Južnej a Severnej pologuli, čo zabezpečuje väčšie pokrytie, avšak nie je najvhodnejšie pre tréningovanie predikčného modelu kvôli čomu bolo potrebné dáta rozdeliť.
- **Vybrať vhodné metriky pre porovnanie predikovaných a pôvodných dát** – výber metrík bol dôležitý z hľadiska nie len výsledného číselného vyjadrenia výsledkov, ale aj ako nástroj pre vyhodnotenie najlepších nastavení modelu.
- **Nájsť vhodné nastavenia vybraného modelu** – vybraný model mal množstvo dostupných nastavení ako pre tréningovanie tak pre vytváranie predikcie.

Bolo, teda potrebné identifikovať nastavenia, ktoré by mohli mať pre našu prácu význam a otestovať ich, aby sme zistili, ktoré sú najvhodnejšie.

- **Natrénovať 2 predikčné modely** – po výbere všetkých vhodných častí bolo potrebné vytvoriť modely pre Južnú a Severnú pologuľu zvlášť pri väčšom množstve dát, z ktorých následne vznikli výsledky našej práce.
- **Vyhodnotiť výsledky predikcie** – po natrénovaní výsledných modelov a ich predikcii bolo potrebné číselne vyjadriť presnosť pomocou vybraných metrík a vyvodiť závery.

3.3 Pochopenie dát

Dáta tejto práce pochádzajú z NASA misie GOLD, ktorá je bližšie opísaná v kapitole 1.2. Pre náš model budeme potrebovať dáta vo forme obrázkov, dostupné sú však vo forme NetCDF súborov. Ide o súbory s koncovým označením `.nc` a tieto súbory obsahujú viac dimenzionálne dáta s viacerými atribútmi. Prvým krokom bolo naučiť sa pracovať s týmto typom súboru. Súbory s dátami sme stiahli zo samotnej stránky misie GOLD, kde sú voľne dostupné pre stiahnutie. Na stránke je taktiež možnosť zobrazenia dát vo forme grafov, avšak vo väčšine prípadov sú zobrazené spolu s inými grafmi, čo nie je možné použiť pre učenie vybraného modelu, a teda si musíme potrebné dáta vo forme grafov vytvoriť. Pre úvodné preskúvanie sme zo stránky stiahli dáta z jedného týždňa, a to konkrétne z 11.10.2021 až 18.10.2021. Pri tejto práci sa zameriavame len na nočné dáta, pretože je na nich najlepšie airglow viditeľný a je taktiež zaujímavý pre jeho vývoj a posun ďalej od hranice slnečného žiarenia (západu slnka z pohľadu povrchu Zeme).

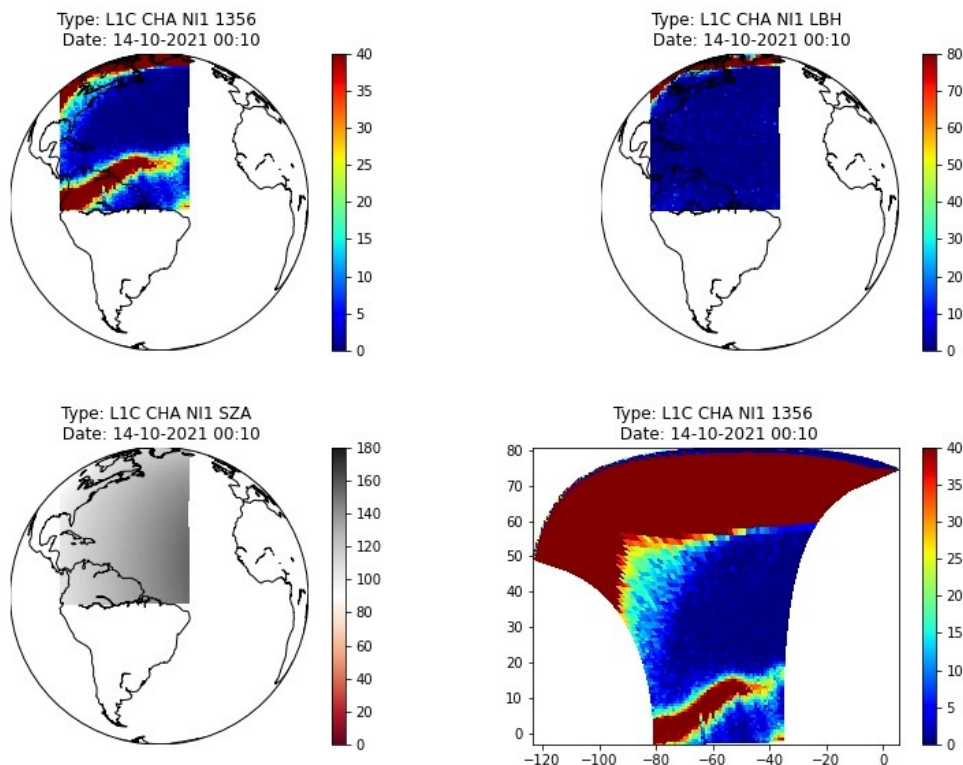
Prvou úlohou pri pochopení dát bolo, čo najlepšie zopakovať zobrazenie, ktoré je možné vidieť na stránke misie GOLD pod možnosťou `quicklook`. Pre prístup k týmto dátam sme použili knižnicu `netCDF4`, ktorá nám umožní pracovať s jednotlivými NetCDF a umožní nám taktiež pristupovať ku jednotlivým atribútom súboru samostatne, teda môžeme si zo súborov vybrať len potrebné atribúty a môžeme ig-

Tabuľka 3 – 1 Zoznam premenných v dátach z misie GOLD ktoré boli dôležité pri tvorbe grafického zobrazenia

Názov	Originálny názov	Jednotky	Popis
Zemepisná dĺžka	Reference_Point_Lon	Stupne	Zemepisná dĺžka referenčného bodu
Zemepisná šírka	Reference_Point_Lat	Stupne	Zemepisná šírka referenčného bodu
Vlnová dĺžka	Wavelength	nm	Vektor vlnových dĺžok pre každý priestorový pixel
Uhol slnečného žiarenia	Solar_Zenith_Angle	Stupne	Rovinný uhol medzi smerom slnka k referenčnému bodu
Intenzita airglow-u	Radiance	Rayleighs/nm	Spektrálne žiarenie v každej polohe mriežky

norovať pre našu prácu nepotrebné atribúty. Z týchto súborov sme následne uložili, ako maskované polia hodnoty zemepisnej dĺžky, zemepisnej šírky, hodnoty žiarenia a vlnovú dĺžku. Detaily o týchto premenných je možné vidieť v Tabuľke 3 – 1. Z údajov vlnovej dĺžky sme následne vybrali len údaje v rozmedzí 135 nm až 137 nm, teda oblasť vlnovej dĺžky označovanej, ako 1356 pričom ide o oblasť atómového kyslíka (135.6 nm). Tieto údaje sme následne premietli ako typ grafu `Pcolor` na zobrazenie Zeme pomocou knižnice `Basemap`. K zobrazeniu sme taktiež pridali farebnú škálu. Pri vlnovej dĺžke 1356 sme taktiež vytvorili graf, ktorý nebol zobrazený na povrchu Zeme, ale išlo len o vykreslenie hodnôt explicitne vo forme grafu `Pcolor`. Pre lepšie pochopenie dát sme taktiež na modeli Zeme vykreslili hodnoty vlnovej dĺžky v rozmedzí 141 nm až 153 nm, čo predstavuje vlnovú dĺžku N2 LBH. Dôležitým faktorom pri nočných dátach je taktiež uhol, pod ktorým na zem dopadajú slnečné lúče. Tieto hodnoty sú v dátach označované ako *SZA* (z angl. *Solar Zenith Angle*). Tieto

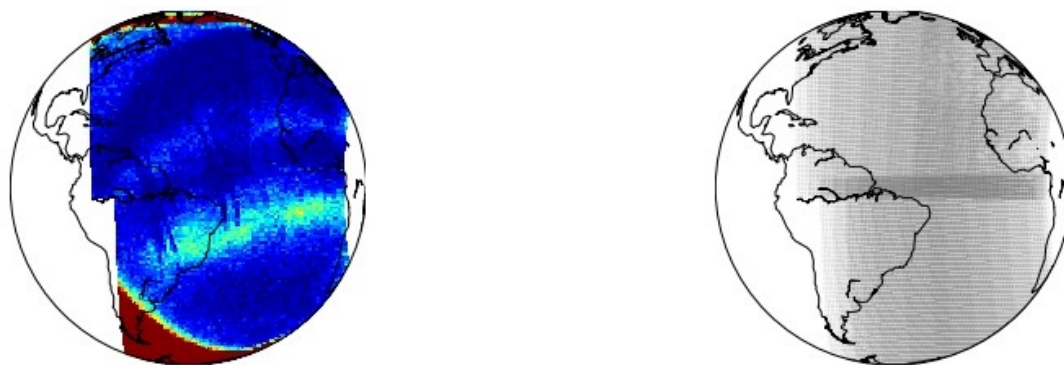
hodnoty sme podobne, ako vlnovú dĺžku uložili, ako maskované pole a zobrazili na povrchu Zeme rovnako, ako pri zobrazení vlnovej dĺžky atómového kyslíka. Pri všetkých týchto grafoch sme taktiež pridali mierku napravo od grafu. Ukážky týchto zobrazení je možné vidieť na Obrázku 3–1.



Obrázok 3–1 Výsledky grafického zobrazenia dát. Prvý obrázok zobrazuje žiarenie 1356. Druhý zobrazuje žiarenie LBH. Na treťom obrázku je možné vidieť SZA. Posledný obrázok znázorňuje celé zobrazenie pre žiarenie 1356 bez podkladu Zeme.

Následne bolo potrebné zistiť akú veľkú časť Zemského povrchu tieto dáta pokrývajú. Pre zistenie tohto faktu sme vytvorili graf, na ktorom sa nachádzali všetky dostupné dáta vlnovej dĺžky 1356 za jednu noc naraz. Tento graf sme vytvorili vďaka spojeniu `Pcolor` grafov pre každý súbor z danej noci, ktoré sme následne zobrazili na tej istej ploche Zeme pomocou knižnice `Basemap`. Na výslednom grafe, ktorý je možné vidieť na Obrázku 3–2 sa nám, teda zobrazila celá plocha pokrytia. Toto zobrazenie sme taktiež vytvorili v priehľadnej sivej variante, vďaka ktorej sme zistili, že

dáta sa najviac prekrývajú v oblasti rovníka. S týmito zisteniami a nájdení spôsobu vytvárania grafických zobrazení dát sme mohli začať pracovať s väčším množstvom dát ako jeden týždeň.



Obrázok 3 – 2 Grafické zobrazenie prekrývania dát

3.4 Príprava dát

Prvou úlohou pri postupe tvorenia tejto práce bolo získanie väčšieho množstva dát, ktoré by sme mohli použiť na tréovanie ako aj testovanie vybraného modelu pre predikciu. Keďže je pre náš výskum potrebné väčšie množstvo dát je pre nás nevhodne použiť stránke misie GOLD, ktorá ponúka grafické rozhranie pre sťahovanie no nemá funkcionality výberu iba nočných dát. Nočné dáta tvoria len zhruba jednu tretinu dát získaných za jeden deň, teda pri použití stránky misie by sme mali veľké množstvo dát, navyše ktoré by bolo potrebné vo výsledku odstrániť. Dáta sme, teda v tomto prípade sťahovali z permanentného archívu SPDF (z angl. *NASA's Space Physics Data Facility*). Na stiahnutie dát sme vytvorili skript v programovacom jazyku Python, kde sme pomocou knižnice `Urllib` analyzovali stránku SPDF a dočasne uložili odkazy podstránky s požadovanými súbormi v požadovanom časom období (dáta za celý rok 2021). Z týchto uložených odkazov sme taktiež odstránili všetky, ktoré neboli zachytené v noci, čo znamená, že neobsahovali v názve súboru `NI1` čo je označenie pre dáta z noci. Následne sme pomocou knižnice `Glob` získali odkazy na každý súbor končiaci na `.nc` v týchto podstránkach pomocou našich ulo-

žených odkazov. Z názvov súborov sme vytvorili priečinky pomenované podľa čísla dňa, v ktorí boli dáta získané. Súborý sme následne stiahli a uložili podľa ich do priečinku s názvom `data` podľa dňa kedy boli získané satelitom misie GOLD. Výsledný priečinkok obsahoval 8 700 NetCDF súborov a celková veľkosť priečinka bola 494 GB. Totožný spôsob bol urobení pre stiahnutie súborov určených na vytvorenie testovacej množiny dát. Tieto dáta musia byť mimo dát použitých na učenie modelu, a teda sme stiahli dáta z prvých 20 dní roku 2020.

Zo stiahnutých súborov bolo následne potrebné vytvoríť dáta v grafickej podobe, ktoré použijeme na tréovanie množiny. Pre vytvorenie grafov sme použili rovnaký proces aký bol opísaný pri pochopení dát, v časti 3.3 ale pracovali sme len s dátami vlnovej dĺžke 1356 a vynechali sme mierku na pravej strane od grafu z dôvodu, aby sa na obrázkoch nachádzal len samotný graf. Dôvodom pre toto rozhodnutie je, aby sme, čo najviac znížili šancu, že model strojového učenia sa pri vytváraní modelu zameria na zlú časť dát. Pri ukladaní grafov sme použili priemer geografickej šírky jednotlivých súborov vďaka čomu bolo možné roztriediť dáta na dve množiny, a to dáta pre Severnú pologuľu a Južnú pologuľu. Tento proces sme opakovali pri všetkých stiahnutých súboroch okrem tých, ktoré boli namerané po polnoci daného dňa, čo spôsobovalo, že ich pri tréovaní ukladalo pred dáta namerané skôr, čo je problémové keďže nami vybraný model sa na dáta pozerá ako na sekvenciu. Dáta z roku 2021 sme pridali do množiny pre tréovanie zatiaľ, čo dáta z roku 2020 do množiny na testovanie. Súbor s dátami sme umiestnili do priečinku samotného modelu Pix2PixHD pod názvom `datasets`.

3.5 Explorácia algoritmu

Keďže sme pripravili dve množiny dát pre každú pologuľu mohli sme sa presunúť k samotnému modelovaniu. Ako model sme si zvolili upravená verziu Pix2PixHD. Jej úprava spočívala v prispôbení modelu na predpovedanie snímok vo videu, teda sekvenciu na seba nadväzujúcich obrázkov. Aj napriek tomu, že nepracujeme

s videom je táto metóda vhodná pre naše dáta keďže snímky airglow-u sú zachytené v časovej postupnosti, a teda na seba nadväzujú. Avšak keďže je tento model určený pre video, obsahuje niektoré nastavenia, ktoré môžeme pre náš model vynechať. Predpovedané výsledky tohto modelu sú tvorené ako nasledujúce snímky sekvencie, čo znamená, že pri vytváraní výsledkov je potrebné mať sekvenciu niekoľkých reálnych snímkov, podľa ktorých bude model vedieť predpovedať snímky, ktoré by mali nasledovať.

Prvým krokom pri modelovaní bolo prísť na správny výber nastavení zvoleného modelu a množstva obrázkov v sekvencii pri vytváraní výsledkov modelu. Zistenie správnych nastavení sme urobili vďaka niekoľkým porovnávacím modelom. Pri tvorbe týchto modelov sme použili menšiu vzorku dát (prvé 2 mesiace roku 2021) kvôli nižšiemu času potrebného na tréningovanie. Všetky modeli sme natrénovali na rovnakých dátoch a taktiež sme použili aj rovnaké testovacie snímky pri vytváraní výsledkov. Výslednú predikciu sme porovnali s originálnymi dátami pomocou zvolených metrík MSE a PSNR. Pred porovnávaním výsledkov s pôvodnými dátami bolo potrebné predpovedané, ale aj pôvodné snímky orezať a previesť do formátu PNG. Toto orezanie sme dosiahli pomocou masky s rozmerom podkladu Zeme, ktorú sme vytvorili pomocou knižnice PIL. Výsledky týchto porovnávacích metrík pri modeloch sme následne porovnali medzi sebou. Všetky nasledujúce kroky boli spustené na serveri určenom na dátovú analytiku od Katedry kybernetiky a umelej inteligencie FEI TUKE. Pre prístup na tento server sme použili rozhranie notebooku Jupyterlab.

Základné nastavenia a príkazy pri spúšťaní tréningovania Pix2PixHD modelu boli následovné:

- **!python**

Časť volajúca základný súbor programovacieho jazyka Python pomocou ktorého môžeme v terminály nastaviť, ktorý súbor má zavolať a aké nastavenia k nemu pridať.

- **pix2pixHD/train_video.py**

Odkazovanie sa na súbor, ktorý má za úlohu spustiť a riadiť tréovanie modelu. Pracuje už so snímkami. O rozdeľovanie videa na snímky sa stará iný súbor, ktorý však v našom prípade nie je potrebný.

- **-- name**

Názov, pod ktorým bude tréovaný model uložený. Tento názov sa aplikuje na priečinok, v ktorom sa bude model nachádzať a taktiež slúži na neskoršie odkazovanie na model v prípade potreby dotréovania alebo vytvárania výsledkov modelu.

- **-- dataroot**

Deklarácia miesta, kde sa nachádzajú vstupné údaje. Je možné použiť aj vlastný priečinok, ale väčšinou sa pri Pix2PixHD nachádzajú vstupné dáta v priečinku `Pix2PixHD/datasets`. Vo vnútri priečinku s množinou dát sa musia nachádzať aspoň 2 súbory, a to `train_frames` pre tréovanie modelu a `test_frames` pre vytvorenie výsledkov modelu.

- **-- save_epoch_freq**

Frekvencia s akou sa budú ukladať epochy modelu. Model taktiež vždy ukladá aj najnovšiu epochu, avšak ukladanie epoch priebežne je dobré pri situáciách kedy by sa model pretréoval alebo by nastala nejaká chyba.

- **pix2pixHD/generate_video.py**

Odkaz na súbor, ktorý má za úlohu vytvoriť video z už vytvoreného modelu. Tento súbor sa stará o vytvorenie výsledkov predikčného modelu, teda snímkov videa a následne z nich vytvoriť funkčné video. Pre našu prácu je funkcia videa nepoužívaná, pretože naša množina dát nepochádza z videa a súbor ho nemôže vytvoriť kvôli veľkosti snímkov. Aj napriek tomu, že má súbor `generate_video.py` tento problém stále vytvorí požadované predpovedané obrázky, čo je výsledok, ktorý potrebujeme.

- `-- how_many`

Číselné vyjadrenie koľko snímkov sa má vytvoriť podľa natrénovaného modelu ako pokračovanie sekvencie snímkov určených na predikciu a testovanie.

- `-- which_epoch`

Číselné vyjadrenie, ktorá epocha sa má pri vytváraní predikcie použiť. Keďže ukladáme zvyčajne každú piatu epochu je možné vybrať hodnoty násobku päťky. Možnosťou je taktiež použiť najnovšiu (ang. latest) ak chceme použiť poslednú natrénovanú epochu.

Pre zistenie vhodného množstva snímkov pri vytváraní predpovedaných nasledujúcich snímkov sme natrénovali model na základných nastaveniach bez akékoľvek zmeny pri tvorbe modelu. Jediný rozdiel spočíval v zmene množstva snímkov v priečinku `test_frames`, ktoré sú používané na vytváranie predikcie z natrénovaného modelu. Vybrali sme 2 rôzne množstvá snímkov. Prvým bolo počet snímkov 5 keďže ide o polovicu dostupných snímkov z jedného dňa. Druhým bolo najmenší možný počet snímkov, ktorý Pix2PixHD dovoľuje, teda 2 snímky. Výsledky predikcie porovnané s originálnymi dátami pri 2 snímkach je možné vidieť v Tabuľke 3–3 a výsledky pri 5 snímkach v Tabuľke 3–2.

Tabuľka 3–2 Výsledky metrík 5 nasledujúcich snímkov pri použití sekvencie 5 snímkov pri vytváraní výsledkov modelu

Čas	MSE	PSNR
22:55	678.48	19.82
23:10	636.03	20.1
23:25	588.27	20.44
23:40	412.63	21.98
23:55	443.95	21.66

Ako je možné vidieť v Tabuľkách 3–2 a 3–3 hodnoty metriky MSE pri 5 snímkach sa približujú k nule omnoho viac, ako pri 2 snímkach a hodnoty PSNR sú

Tabuľka 3–3 Výsledky metrík 5 nasledujúcich snímok pri použití sekvencie 2 snímok pri vytváraní výsledkov modelu

Čas	MSE	PSNR
22:55	3134.09	13.17
23:10	3449.1	12.75
23:25	3099.92	13.22
23:40	2543.58	14.08
23:55	2161.66	14.78

taktiež väčšie pri 5 snímkach. Tieto hodnoty sú zapríčinené tým, že model je pôvodne určený na video, ktoré by malo omnoho väčšie množstvo snímok. Teda platí čím je väčší počet snímok tým je väčšia aj presnosť predikcie. V práci sme sa ďalej rozhodli pracovať so sekvenciou 5 snímok z dôvodu, že chceme, aby model vytvoril väčšiu sekvenciu obrázkov, aj keď najväčšiu presnosť by sme dosiahli pri predpovedi len posledného snímku z danej noci.

Pre zistenie vhodných nastavení algoritmu sme po dohode s konzultantmi identifikovali štyri nastavenia, ktoré by mohli mať dopad na výsledný model alebo sme si pri nich neboli istý aký dopad by mohli mať na model, a teda ich bolo potrebné otestovať. Pre každé z týchto nastavení sme následne natrénovali model pri menšom počte dát pričom ostatné nastavenia sme nechali pôvodné. Výsledky modelov boli následne vytvorené ako pokračovanie sekvencie 5 snímok z danej noci. Vybrané nastavenia boli nasledujúce:

- **-- batchsize**

Ide o nastavenie počtu tréningových príkladov použitých v jednej iterácii tréningovania modelu. Pri tomto nastavení je základná hodnota 1 pričom pri testovaní sme nastavili hodnotu na 5. Výsledky tréningovania s nastavením `-- batchsize` porovnané s originálnou množinou dát je možné vidieť v Tabuľke 3–4.

- **-- fps**

Tabuľka 3–4 Výsledky metrík 5 nasledujúcich snímkov pri použití nastavenia `-- batchsize 5`.

Čas	MSE	PSNR
22:55	681.60	19.8
23:10	710.18	19.62
23:25	720.65	19.55
23:40	518.5	20.98
23:55	510.19	21.05

Nastavenie určené pre predpovedanie snímkov vo videu. Označuje koľko snímkov za sekundu sa vo videu nachádza, teda aj aká je frekvencia vstupných dát. Keďže naše dáta nepochádzajú z videa bolo potrebné vyskúšať aký vplyv toto nastavenie má, a tak sme ho skúsili nastaviť na hodnotu 1. Výsledky tréningovania s nastavením `-- fps` porovnané s originálnou množinou dát je možné vidieť v Tabuľke 3–5.

Tabuľka 3–5 Výsledky metrík 5 nasledujúcich snímkov pri použití nastavenia `-- fps 1`.

Čas	MSE	PSNR
22:55	668.25	19.88
23:10	599.86	20.35
23:25	592.24	20.41
23:40	369.02	22.46
23:55	423.87	21.86

- `-- no_flip`

Ak je toto nastavenie použité, vstupné obrázky sa neprevracajú kvôli zväčšeniu množstva údajov. Výsledky tréningovania s nastavením `-- no_flip` porovnané s originálnou množinou dát je možné vidieť v Tabuľke 3–6.

Výsledkov metrík pri zmene jednotlivých nastavení z Tabuliek 3–4, 3–5 a 3–6 sme porovnali s výsledkami z Tabuľky 3–2 kde boli použité základne nastavenia pri

Tabuľka 3–6 Výsledky metrík 5 nasledujúcich snímkov pri použití nastavenia `-- no_flip`.

Čas	MSE	PSNR
22:55	648.01	20.01
23:10	636.91	20.09
23:25	616.32	20.23
23:40	416.06	21.94
23:55	478.57	21.33

rovnako dlhej sekvencii obrázkov. Z tabuliek sme zistili, že zmeny medzi jednotlivými nastaveniami boli minimálne, a tak sme sa rozhodli pokračovať v modelovaní so základnými nastaveniami modelu Pix2PixHD.

3.6 Výsledné modely a vyhodnotenie

Po zistení všetkých vhodných nastavení a vhodného množstva obrázkov v sekvencii pri vytváraní výsledkov sme následne natrénovali dva modeli na všetkých dátach, ktoré sme mali určené pre tréning (všetky dáta za rok 2021). Dôvod prečo sme sa rozhodli vytvoriť 2 modeli miesto jedného je, že dáta sú snímané striedavo (snímok zo Severnej pologule a následne Južnej pologule). Takáto „skáková“ sekvencia snímok by spôsobovala väčšiu nepresnosť pri učení, ako pri rozdelených modeloch. Oba tieto modely boli natréňované až po epochu 200, pričom sme ukladali každú piatu epochu. Celé tréningové týchto modelov zabralo približne 5 dní nepretržitého tréningu na servery od Katedry kybernetiky a umelej inteligencie FEI TUKE. Vďaka použitiu servera mohlo tréningové prebiehať 24 hodín denne, čo by na vlastných zariadeniach nebolo možné.

Výsledné hodnoty vybraných metrík pri použití na výsledky natréňovaný modelov je možné vidieť v Tabuľke 3–7. Pri metrike MSE platí, čím je hodnota bližšie k nule tým je snímok podobnejší k tomu očakávanému (pôvodnému). Pri metrike PSNR platí, že čím je číslo väčšie tým sú obrázky podobnejšie. Pri rovnakých ob-

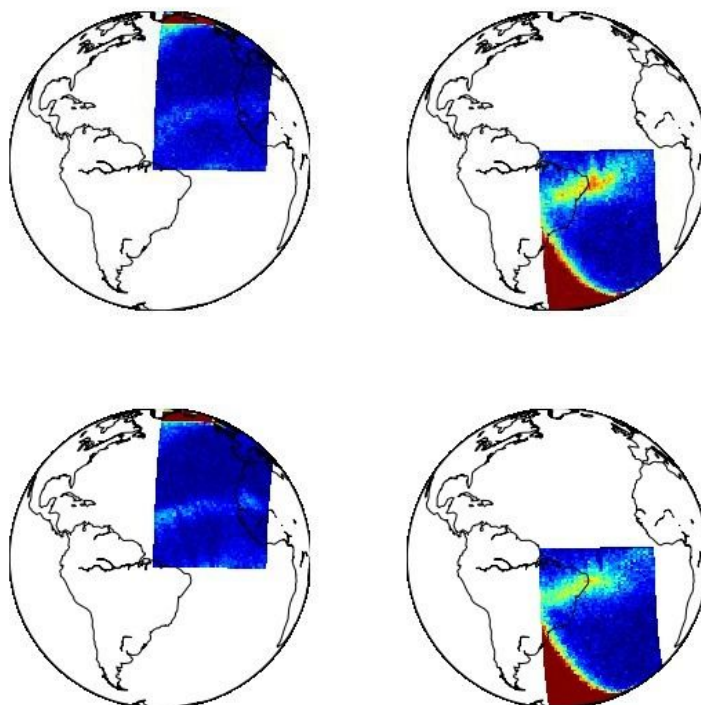
rázkoch sa však nedá hovoriť o presnej hodnote, ako pri MSE, pretože, ako je možné vidieť v sekcii 2.4 pri vzorci PSNR sa používa odmocnina metriky MSE, ktorá je pri rovnakých snímkach rovná 0 čím sa deliť nedá. V Tabuľke 3–7 môžeme vidieť hodnoty pre niekoľko vybraných dní z Januára 2020. Dáta z 20. Januára sú použité z dôvodu, že sa počas tohto dňa na dátach nenachádzal žiaden výrazný úkaz. Hodnoty metrík sú pri tomto dni o niečo lepšie, ako pri ostatných, čo je spôsobené tým, že snímky sú takmer kompletne celé len modrej farby, a teda predikcia bola jednoduchšia a je aj presnejšia. Tento fakt je taktiež úspechom, pretože náš model pri tejto predikcii nevytvoril falošne pozitívne snímky. Pri dátach z 5. Januára sa na Severnej aj Južnej pologuli už nachádzali úkazy, čo je viditeľné, aj na hodnotách metrík, ktoré sú horšie, ako pri dátach z 20. Januára, pretože predikcia už nebola taká priamočiara. Metriky sú stále však v prijateľných hodnotách, avšak nemáme ich ako porovnať s inými výsledkami keďže metódy hlbokého učenia na nami používané dáta ešte neboli otestované. Dáta z 2. Januára boli použité iba pri severnom modeli a z 10. Januára iba pri južnom modeli, a to z dôvodu, že pri týchto dňoch na opačnej pologuli neboli žiadne výrazné úkazy, ktoré by bolo možné pozorovať. Hodnoty metrík sa v týchto 2 dňoch výrazne líšia, a to aj z dôvodu, že na Južnej pologuli sa počas roku vyskytuje menšie množstvo výrazných úkazov, a teda, aj model pre južnú pologulu ich nevie odhadovať s takou presnosťou, ako pri modeli Severnej pologule. Prvú predikciu z týchto dvoch dní je taktiež možné vidieť na obrázku 3–3.

Na Obrázku 3–4 je možné vidieť kompletnú sekvenciu, ktorá bola vytvorená pomocou modelu pre Severnú pologulu z dňa 20.1.2020. Išlo o pokračovanie sekvencie 5 snímkov. Za noc je misiou GOLD zachytených snímkov 10, teda predpovedáme polovicu noci. Predpovedaných snímkov je však 6 z dôvodu, že algoritmus Pix2PixHD predpovedá aj posledný snímok z predom vlozenej sekvencie (v našom prípade piaty snímok z noci). Na obrázku je možné vidieť, že najpresnejší je hneď prvý snímok pričom pri zvyšku sekvencie už je menej presný no stále si udržuje celkom dobrú presnosť predpovedania svetlejšie modrej oblasti, kde sa nachádza slabšie žiarenie airglow-u. Táto presnosť, ktorú vidíme taktiež zodpovedá metrikám v Tabuľke 3–7.

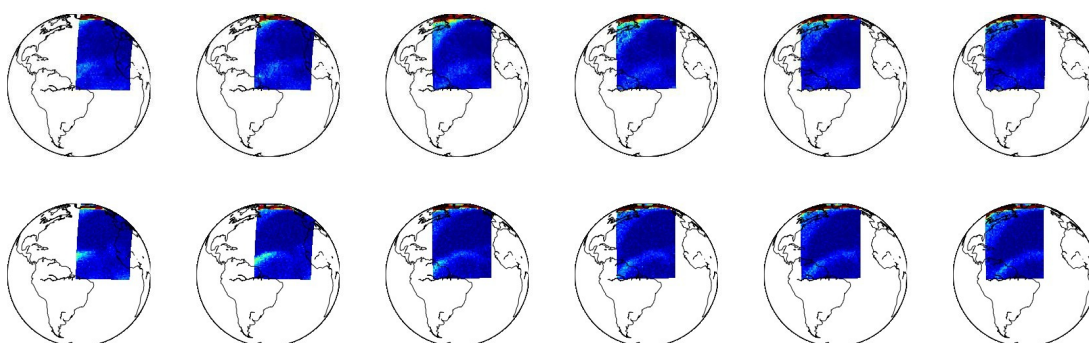
Tabuľka 3–7 Výsledky metrik 6 po sebe idúcich snímok pre Severnú a Južnú pologuľu pri rôznych vybraných dňoch Januára 2020

Severná pologuľa			Južná pologuľa		
Čas	MSE	PSNR	Čas	MSE	PSNR
02.01.2020			10.01.2020		
22:10	83.26	28.93	22:25	170.15	25.82
22:40	141.87	26.61	22:55	324.03	23.02
23:10	197.96	25.16	23:10	315.13	23.15
23:25	242.04	24.29	23:25	311.69	23.19
23:40	180.05	25.58	23:40	374.3	22.4
23:55	170.6	25.81	23:55	348.61	22.71
05.01.2020			05.01.2020		
22:10	129.66	27.01	22:25	117.19	27.44
22:40	172.58	25.76	22:55	182.18	25.53
23:10	247.51	24.19	23:10	138.88	26.7
23:25	350.16	22.69	23:25	137.61	26.74
23:40	355.01	22.63	23:40	217.11	24.76
23:55	366.92	22.49	23:55	210.99	24.89
20.01.2020			20.01.2020		
22:10	94.53	28.37	22:25	93.26	28.43
22:40	168.9	25.85	22:55	159.15	26.11
23:10	183.63	25.49	23:10	145.11	26.51
23:25	224.43	24.62	23:25	128.97	27.03
23:40	131.59	26.94	23:40	220.26	24.7
23:55	204.29	25.03	23:55	184.03	25.48

Cieľ tejto práce bolo vytvorenie modelu airglow-u pomocou hlbokého učenia. Aby sme sa k tomuto výsledku dostali bolo počas tejto práce potrebné oboznámiť



Obrázok 3–3 Porovnanie predpovedaných (horná časť) a reálnych dát (dolná časť) našich modelov.



Obrázok 3–4 Porovnanie predpovedanej (horná časť) a reálnej (dolná časť) sekvencie severného modelu pre 20.1.2020

sa hlavne s dátami misie NASA GOLD a ich spracovaním. To sa nám úspešne podarilo vytvorením množiny dát v podobe snímok. Následne bolo potrebné sa oboznámiť s vybraným modelom hlbokého učenia Pix2PixHD a zistiť jeho vhodné

nastavenia. Cieľ sa nám podarilo splniť, pretože sme vytvorili 2 modeli pre predikciu airglow, ktorých presnosť je dostačujúca, avšak nie úplne najlepšia. Keďže ide o prvý pokus využitia modelu hlbokého učenia na tieto dáta nie je možné ich porovnať s inými výsledkami. V práci sa nachádza niekoľko faktorov, ktoré by bolo možné v budúcnosti vylepšiť alebo lepšie preskúmať. Jedným z nich sú hodnoty airglow-u na okrajoch Zemského disku, ktoré sme v tejto práci nebrali do úvahy. Pridanie tejto oblasti do snímok by mohlo rozšíriť plochu predikcie jednotlivých snímok. Ďalším bodom možného vylepšenia by mohlo byť otestovanie viacerých algoritmov, a to aj tých, ktoré by pracovali s pôvodnými dátami a nie so snímkami. Takéto Algoritmy by mohli mať potencionálne lepšiu presnosť no boli by omnoho zložitejšie.

4 Záver

V tejto práci sme sa venovali airglow-u pričom sme opísali miesto jeho vzniku a pozorovanie tohto javu. Venovali sme sa taktiež NASA misii GOLD, v ktorej sme po teoretickej časti opísali jej snímanie airglow-u a dáta, ktoré poskytuje. Pre prácu s týmito dátami sme vybrali algoritmus Pix2PixHD, ktorý je pôvodne určený na prácu s videom, no my sme ho použili pri práci so sekvenciou snímok z misie GOLD. Dáta misie sme upravili do grafickej podoby a po nájdení vhodných nastavení vybraného algoritmu sme vytvorili predikčné modely pre Severnú a Južnú pologuľu.

Výsledky týchto modelov sme číselne vyjadrili pomocou metrík MSE a PSNR pričom hodnoty, ktoré sme dostali boli dostačujúce vzhľadom na to, že neexistujú žiadne vydané práce, ktoré by pracovali s predikovaním týchto dát. Tento fakt taktiež spôsobil, že sme výsledné hodnoty metrík nášho riešenia nemali ako porovnať s inými výsledkami. Plán pre túto prácu je taktiež ju v budúcnosti ešte rozšíriť a publikovať výsledky v odbornom časopise.

Pre vylepšenie našich výsledkov bude v budúcnosti možné použiť lepšie grafické zobrazenie dát, ktoré bude zahŕňať aj oblasti na okraji disku. Možnosťou pre vylepšenie a ďalšie skúmanie by taktiež mohlo byť aplikovať na dáta iné modely, a to aj tie, ktoré budú pracovať s číselnými dátami na rozdiel od grafických.

Oblasť skúmania airglow-u je stále vyvíjajúca sa činnosť no použitie hlbokého a strojového učenia má v tejto oblasti veľký potenciál a môže priniesť zaujímavé výsledky, ktoré prispejú k pochopeniu airglow-u samotnému ako aj oblasti hornej atmosféry Zeme.

Literatúra

- Butler, K. (2018). Accessing multidimensional scientific data using python. <https://www.esri.com/arcgis-blog/products/arcgis-pro/analytics/accessing-multidimensional-scientific-data-using-python/>.
- Case, N. (2021). Beautiful green 'airglow' spotted by aurora hunters – but what is it?, *The Conversation* . <https://theconversation.com/beautiful-green-airglow-spotted-by-aurora-hunters-but-what-is-it-68188>.
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. and Bharath, A. A. (2018). Generative adversarial networks: An overview, *IEEE Signal Processing Magazine* **35**(1): 53–65.
- Demir, U. and Ünal, G. B. (2018). Patch-based image inpainting with generative adversarial networks, *CoRR* [abs/1803.07422](https://arxiv.org/abs/1803.07422). <http://arxiv.org/abs/1803.07422>.
- Deng, L. (2012). The mnist database of handwritten digit images for machine learning research [best of the web], *IEEE Signal Processing Magazine* **29**(6): 141–142.
- Eastes, R. W., McClintock, W. E. and et al. (2017). The global-scale observations of the limb and disk (gold) mission, *Space Science Reviews* **212**(1): 383–408. <https://doi.org/10.1007/s11214-017-0392-2>.
- Eastes, R. W., McClintock, W. E. and et al. (2020). Initial observations by the gold mission, *Journal of Geophysical Research: Space Physics* **125**(7): e2020JA027823. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JA027823>.
- Gold Instrument* (n.d.). <https://gold.cs.ucf.edu/science-mission/gold-instrument/>.

-
- Hahne, F., Huber, W., Gentleman, R. and Falcon, S. (2008). *Bioconductor case studies*, Springer. <https://link.springer.com/book/10.1007/978-0-387-77240-0>.
- Harris, C. R., Millman, K. J. and et al. (2020). Array programming with numpy, *Nature* **585**(7825): 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
- HECHT-NIELSEN, R. (1992). Iii.3 - theory of the backpropagation neural network**based on “nonindent” by robert hecht-nielsen, which appeared in proceedings of the international joint conference on neural networks 1, 593–611, june 1989. © 1989 ieee., pp. 65–93. <https://www.sciencedirect.com/science/article/pii/B9780127412528500108>.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment, *Computing in Science Engineering* **9**(3): 90–95.
- Kaelbling, L. P., Littman, M. L. and Moore, A. W. (1996). Reinforcement learning: A survey, *J. Artif. Int. Res.* **4**(1): 237–285.
- Lai, C., Xu, J., Yue, J., Yuan, W., Liu, X., Li, W. and Li, Q. (2019). Automatic extraction of gravity waves from all-sky airglow image based on machine learning, *Remote Sensing* **11**(13). <https://www.mdpi.com/2072-4292/11/13/1516>.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning, *Nature* **521**(7553): 436–444. <https://doi.org/10.1038/nature14539>.
- Lewin, S. (2018). Nasa’s gold mission powers on for first time after eventful launch. <https://www.space.com/39547-nasa-gold-powers-on-first-time.html>, journal=Space.com.
- Lühr, H. and Rother, M. (2003). The diamagnetic effect of the equatorial appleton anomaly: Its characteristics and impact on geomagnetic field modeling, *Geophysical Research Letters* **30**(17). <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GL017407>.
-

-
- Mackovjak, , Varga, M., Hrivňak, S., Palkoci, O. and Didebulidze, G. G. (2021). Data-driven modeling of atomic oxygen airglow over a period of three solar cycles, *Journal of Geophysical Research: Space Physics* **126**(3): e2020JA028991. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JA028991>.
- Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets, *CoRR abs/1411.1784*. <http://arxiv.org/abs/1411.1784>.
- Quick facts: Global-scale observations of the limb and disk (gold)* (2019). <https://lasp.colorado.edu/home/missions-projects/quick-facts-gold/>.
- Raschka, S. (2015). *Python Machine Learning*, Packt Publishing. <https://books.google.sk/books?id=GOVOCwAAQBAJ>.
- Ren, X., Li, H., Huang, Z. and Chen, Q. (2020). *Self-Supervised Dance Video Synthesis Conditioned on Music*, Association for Computing Machinery, New York, NY, USA, p. 46–54. <https://doi.org/10.1145/3394171.3413932>.
- Ronneberger, O., Fischer, P. and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, pp. 234–241.
- Russo, A. and Lax, G. (2022). Using artificial intelligence for space challenges: A survey, *Applied Sciences* **12**: 5106.
- Samuel, A. L. (2000). Some studies in machine learning using the game of checkers, *IBM Journal of Research and Development* **44**(1.2): 206–226. <https://doi.org/10.1147/rd.441.0206>.
- Savigny, C. v. (2017). Airglow in the earth atmosphere: basic characteristics and excitation mechanisms, *ChemTexts* **3**(4): 14. <https://doi.org/10.1007/s40828-017-0051-y>.

SES-14 Integrates NASA Ultraviolet Space Spectrograph

-
- (2017). *SES*. <https://www.ses.com/press-release/ses-14-integrates-nasa-ultraviolet-space-spectrograph>.
- Silva, T. (2020). An intuitive introduction to generative adversarial networks (gans). <https://www.freecodecamp.org/news/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394>.
- Singh, A., Thakur, N. and Sharma, A. (2016). A review of supervised machine learning algorithms, pp. 1310–1315.
- Tran, L. (2018). Why nasa watches the colors of the (upper atmospheric) wind, *NASA*. <https://www.nasa.gov/feature/goddard/2018/why-nasa-watches-airglow-the-colors-of-the-upper-atmospheric-wind>.
- Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE, *Mind* **LIX**(236): 433–460. <https://doi.org/10.1093/mind/LIX.236.433>.
- Wang, T., Liu, M., Zhu, J., Tao, A., Kautz, J. and Catanzaro, B. (2017). High-resolution image synthesis and semantic manipulation with conditional gans, *CoRR* **abs/1711.11585**. <http://arxiv.org/abs/1711.11585>.
- Wang, Z. and Bovik, A. C. (2009). Mean squared error: Love it or leave it? a new look at signal fidelity measures, *IEEE Signal Processing Magazine* **26**(1): 98–117.
- Wang, Z., Bovik, A., Sheikh, H. and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing* **13**(4): 600–612.
- Whitaker, J. (2011). <https://matplotlib.org/basemap/users/intro.html>.

Zoznam príloh

Príloha A Systémová príručka

Príloha B Používateľská príručka

Príloha C CD médium obsahujúce všetky zdrojové kódy, bakalársku prácu a prílohy v elektronickej podobe