ՀՀ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԶԳԱՅԻՆ ԱԿԱԴԵՄԻԱ
ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

Պողոսյան Աղասի Ստեփանի

ԹՎԱՅԻՆ ՊԱՏԿԵՐՆԵՐԻ ԻՄԱՍՏԱԲԱՆԱԿԱՆ ՎԵՐԼՈՒԾՈՒԹՅԱՆ ՄԵԹՈԴՆԵՐԻ ՄՇԱԿՈՒՄ ԵՎ ԻՐԱԿԱՆԱՑՈՒՄ

Ե.13.05. – «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի համալիրներ» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի համար

Ս Ե Ղ Մ Ա Գ Ի Ր

Երևան – 2019

ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ НАН РА

Погосян Агаси Степанович

РАЗРАБОТКА И РЕАЛИЗАЦИЯ МЕТОДОВ СЕМАНТИЧЕСКОГО АНАЛИЗА ЦИФРОВОГО ИЗОБРАЖЕНИЯ

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени кандидата технических наук по специальности 05.13.05 «Математическое моделирование, численные методы и комплексы программ»

Ереван – 2019

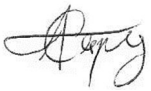
© National Library of Armenia OCR by Portmind

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում:

Գիտական ղեկավար՝	ֆիզմաթ. գիտ. դոկտոր	Հ. Գ. Սարուխանյան
Պաշտոնական ընդդիմախոսներ՝	ֆիզմաթ. գիտ. դոկտոր	Է. Մ. Պողոսյան
	ֆիզմաթ. գիտ. թեկնածու	Ա. Վ. Պողոսյան

Առաջատար կազմակերպություն՝ Հայաստանի ազգային պոլիտեխնիկական համալսարան

Պաշտպանությունը կայանալու է 2019թ. հունիսի 14-ին, ժ. 17:00-ին ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում գործող 037 «Ինֆորմատիկա» մասնագիտական խորհրդի նիստում, հետևյալ հասցեով՝ Երևան, 0014, Պ. Սևակի 1:
Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ ԻԱՊԻ գրադարանում:
Սեղմագիրը առաքված է 2019թ. մայիսի 3-ին:

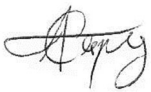
037 Մասնագիտական խորհրդի
գիտական քարտուղար ֆ.մ.գ.դ.  Հ.Գ.Սարուխանյան

Тема диссертации утверждена в Институте проблем информатики и автоматизации НАН РА.

Научный руководитель:	доктор физ.-мат.наук	А.Г.Саруханян
Официальные оппоненты:	доктор физ.-мат. наук	Э. М. Погосян
	кандидат физ.-мат. наук	А. В. Погосян

Ведущая организация: Национальный политехнический университет Армении

Защита состоится 14-ого июня 2019г. в 17:00 на заседании специализированного совета 037 «Информатика» Института проблем информатики и автоматизации НАН РА по адресу: 0014, г. Ереван, ул. П. Севака 1.
С диссертацией можно ознакомиться в библиотеке ИПИА НАН РА.
Автореферат разослан 3-ого мая 2019г.

Ученый секретарь,
Специализированного совета 037
доктор физ.-мат.наук  А.Г.Саруханян

© National Library of Armenia

ԱՇԽԱՏԱՆՔԻ ԸՆԴՀԱՆՈՒՐ ԲՆՈՒԹԱԳԻՐԸ

Աշխատանքի արդիականությունը: Թվային պատկերների իմաստաբանական վերլուծությունը տեղեկատվական ճարտարագիտության (Information Engineering) արդի խնդիրներից է: Դրա վերջնական լուծումը ենթադրում է, որ կստեղծվի համակարգ, որը կկարողանա նկարագրել պատկերներն այնպես, ինչպես կարողանում է մարդը: Լուծումն առնչվում է այնպիսի ոլորտներին, ինչպիսիք են մեքենայական ուսուցումը (Machine learning), վիճակագրությունը (Statistics), բնական լեզուների մշակումը (Natural language processing), համակարգչային տեսողությունը (Computer vision), թվային ազդանշանների և պատկերների մշակումը (Signal and Image Processing), ինչպես նաև՝ ինֆորմացիայի տեսությունը (Information theory) և նեյրոկենսաբանությունը (Neurobiology):

Թվային պատկերների իմաստաբանական վերլուծություն կատարելը ենթադրում է, որ համակարգի մուտքին տրվող պատկերի համար պետք է ձևավորվեն այնպիսի արդյունքներ, որոնք արտահայտված կլինեն բնական լեզուների տարրերով՝ բառերով և նախադասություններով: Ընդ որում պատկերը նկարագրող առանձին վերցված ֆիզիկական մեծությունները, օրինակ՝ պիքսելի գույնը կամ պատկերի պայծառությունը, չեն կարող համարվել դրա իմաստաբանական գնահատական, սակայն դրանք կարող են հանդիսանալ բառակապակցության կամ նախադասության մաս՝ օրինակ «կարմիր մեքենա», «տղամարդը նստած է պայծառ սենյակում»: Խնդրի հիմնական բարդությունը կայանում է նրանում, որ պատկերների իմաստաբանական վերլուծության մաս կազմող օբյեկտների հայտնաբերումը և տեղորոշումը չի լուծվում միայն պատկերների մշակման և նմուշների ճանաչման մեթոդներով (Pattern recognition), քանի որ պատկերում գտնվող «օբյեկտները»՝ կախված միմյանց նկատմամբ դասավորվածությունից և մասշտաբից, կարող են դիտարկվել որպես տարբեր առարկաներ: Օրինակ՝ առանձին վերցված կենդանու մազերը կարող են շատ նման լինել սարդոստայնների: Այդ պատճառով պատկերների սեգմենտացիայի ալգորիթմները դառնում են անզոր այս խնդրի լուծման համար: Հենց նույն կերպ էլ հնարավոր չի լինում հայտնաբերել օբյեկտները, քանի որ անհայտ է, թե պատկերի որ հատվածները պետք է դրա համար համեմատվեն: Ստացվում է, որ հնարավոր չէ կատարել նմուշների (pattern) համեմատում, քանի որ հայտնի չեն պատկերում օբյեկտների եզրագծերը, իսկ դրանց հայտնաբերման համար էլ անհրաժեշտ է նախապես իմանալ օբյեկտների տեսակները: Այսպիսով, պատկերում օբյեկտների հայտնաբերում և տեղորոշում կատարող համակարգը պետք է օժտված լինի իրարից ուղղակիորեն կախվածություն չունեցող և՛ ճանաչման, և՛ տեղորոշման հատկություններով: Նմանատիպ խնդիրներ ի հայտ են գալիս նաև պատկերը մեկ ընդհանրական վերնագրով (caption) նկարագրելիս, որտեղ առաջանում են բնական լեզվով նախադասության գեներացման հետ կապված խնդիրներ:

Վերջին տարիներին ստեղծվել են պատկերների իմաստաբանական վերլուծություն կատարող տարբեր համակարգեր: Դրանցից են պատկերները դասակարգող խորը փաթույթային նեյրոնային ցանցերի վրա հիմնված համակարգերը: Դրանց կիրառմամբ մշակված են պատկերներում օբյեկտներ հայտնաբերող, ճանաչող և տեղորոշող, ինչպես նաև պատկերների վերնագրերը բնական լեզվով գեներացնող ցանցեր: Այդ ցանցերի վրա հիմնված համակարգերը կիրառվում են ռոբոտաշինության մեջ, օրինակ՝

© National Library of Armenia

ինքնակառավարվող ավտոմեքենաներ և անօդաչու թռչող սարքեր նախագծելիս, բացի դրանից, ճիշտ ներդնելու դեպքում, կունենան պոտենցիալ այլ կիրառություն, օրինակ՝ հիմնովին կարող են հեշտացնել տեսողական խնդիրներ ունեցող մարդկանց առօրյան:

Պատկերների իմաստաբանական վերլուծությունների կարևոր կիրառություններից են նաև որոնող (search) և խորհրդատու (recommendation) համակարգերը, որոնք բանալի բառերով կատարում են պատկերների որոնում: Որոնումը լավ կազմակերպելու նպատակով անհրաժեշտություն է առաջանում հայտնաբերել պատկերում գտնվող օբյեկտները և նկարագրել դրանց հարաբերություններն ու հատկությունները: Հաճախ պատկերների իմաստաբանական վերլուծության համար որոշ լուծումներում համատեղ կիրառվում են տարբեր իմաստաբանական վերլուծություն կատարող համակարգեր: Բացի դրանցից որոշ լուծումներ օգտագործում են դինամիկ օպտիմիզացիայի մեթոդներ, որոնք շատ դանդաղագործ են: Նշենք, որ ժամանակակից խորհրդատու և փնտրող համակարգերում մեծ կարևորություն ունի արագագործությունը, որի պատճառով վերը նշված լուծումներն անարդյունավետ են:

Այսպիսով՝ իրական ժամանակում աշխատող համակարգերում պատկերների իմաստաբանական վերլուծություն կատարող համակարգերը պետք է հնարավորինս քիչ հաշվարկներ կատարելով կարողանան գեներացնել մուտքային պատկերների մանրամասն նկարագրություն:

Ատենախոսության նպատակն է մշակել թվային պատկերների իմաստաբանական վերլուծություն կատարող արագագործ մեթոդ, որը կկարողանա մուտքային պատկերը վերնագրել բնական լեզվի քերականությանը համապատասխանող նախադասությամբ, ճանաչել և տեղորոշել պատկերներում գտնվող օբյեկտները, ինչպես նաև հայտնաբերված օբյեկտների համար հաշվել գերիշխող գույները: Մշակված մեթոդի համար անհրաժեշտ է ստեղծել ծրագրային համակարգ, որը հաշվարկների արագացման համար կօգտագործի համակարգչի թե՛ կենտրոնական և թե՛ գրաֆիկական պրոցեսորների բոլոր միջուկները:

Հետազոտության մեթոդներ: Թվային պատկերների իմաստաբանական վերլուծության մեթոդների մշակման համար օգտագործվել են պատկերները դասակարգող, օբյեկտներ հայտնաբերող և տեղորոշող խորը փաթույթային ու պատկերների վերնագրերը բնական լեզվով արտահայտող ռեկուրենտ նեյրոնային ցանցեր: Մոդելների ուսուցման, գնահատման և թեստավորման համար ստեղծված համակարգերում օգտագործվել են զուգահեռ հաշվարկների մեթոդներ:

Արդյունքների գիտական նորությունները

- Մշակվել է պատկերների վերնագրեր գեներացնող ռեկուրենտ նեյրոնային ցանց, որը հիմնված է օբյեկտներ հայտնաբերող և տեղորոշող խորը փաթույթային նեյրոնային ցանցից դուրս բերված բարձր մակարդակի հատկությունների վրա: Արդյունքում ստեղծվել է պատկերների վերնագրեր գեներացնող, պատկերներում օբյեկտներ հայտնաբերող և տեղորոշող մեկ միասնական, արագագործ ցանց: Ցանցերի ուսուցման, գնահատման և թեստավորման համար ստեղծվել են ծրագրային համակարգեր, որոնք հաշվարկները զուգահեռ կատարում են կենտրոնական և գրաֆիկական

© National Library of Armenia

պրոցեսորների բոլոր միջուկների վրա: Մշակվել է իրական ժամանակում աշխատող ծրագրային համակարգ, որը կատարում է պատկերների վերնագրերի գեներացում, օբյեկտների հայտնաբերում, ճանաչում և տեղորոշում, ինչպես նաև հաշվում է օբյեկտների գերիշխող գույները:

- Մշակվել է իմաստաբանորեն նման պատկերների որոնման ալգորիթմ, որը GoogLeNet ցանցից դուրս բերված բարձր մակարդակի հատկությունների կիրառմամբ կատարում է օբյեկտների որոնում 89% ճշգրտությամբ:
- Մշակվել է մեծ կարճաժամկետ հիշողության (Long Short-Term Memory) մոդիֆիկացված տարբերակ, որը պատկերների վերնագրերի գեներացման խնդրում ցուցաբերել է ավելի բարձր ճշգրտություն: Ուսումնասիրության համար ստեղծվել է ծրագրային համակարգ, որը թույլ է տալիս մոտ 10 անգամ կրճատել ուսուցման մեկ փորձի վրա ծախսվող ժամանակը և նվազեցնել գրաֆիկական քարտի հիշողության օգտագործումը:

Աշխատանքի կիրառական նորությունները: Մշակված պատկերների իմաստաբանական վերլուծություն կատարող համակարգը կարելի է ներդնել որոնող համակարգերում՝ նկարագրությամբ կամ բանալի բառերով պատկերների որոնման համար: Տեսողության խնդիրներ ունեցող մարդկանց կարող է օգնել ինչպես վիրտուալ, այնպես էլ իրական աշխարհում՝ տալով այնտեղ գտնվող առարկաների և իրադարձությունների մասին ինֆորմացիա: Ներդնելով մշակված ծրագրային միջոցները՝ կարելի է ավտոմատացնել որոշ հսկողության համակարգեր: Լայն կիրառություն կարող է ունենալ թե՛ արտադրական, թե՛ այլ նշանակության համակարգերում:

Մշակված իմաստաբանորեն նման պատկերների որոնման ալգորիթմի կիրառմամբ կարելի է խմբավորել (clustering) չպիտակավորված պատկերները, ինչպես նաև խորհրդատու (recommendation) և որոնող (search) համակարգերում կազմակերպել նման պատկերների որոնում:

Ներդրում: Ատենախոսության շրջանակներում ստացված արդյունքները և մշակված միջոցները ներդրվել են «ՊիկսԱրտ ՍՊԸ» ընկերությունում՝ պատկերների իմաստաբանական որոնում և չպիտակավորված պատկերների դասակարգում կազմակերպելու նպատակով: Մշակված մեթոդները կիրառվել են ընկերության սերվերային համակարգերում՝ պատկերի նմանակների հայտնաբերման, ինչպես նաև չպիտակավորված պատկերների պիտակավորման, դասակարգման և խորհրդատու համակարգի կատարելագործման համար:

Աշխատանքի արդյունքները զեկուցվել են.

- «Համակարգչային գիտություններ և տեղեկատվական տեխնոլոգիաներ» միջազգային գիտաժողովում (CSIT 2017, ք. Երևան, Հայաստան),
- «International Conference Dedicated to 90th Anniversary of SERGEY MERGELYAN» գիտաժողովում (2018թ., ք. Երևան, Հայաստան),
- «Գիտության և տեխնոլոգիաների մերձեցում» գիտաժողովում (2018թ., ք. Երևան, Հայաստան),

© National Library of Armenia

- ՀՀ ԳԱԱ ԻԱՊԻ կոդավորման և ազդանշանների մշակման գիտամեթոդական սեմինարում (2018թ.),
- ՀՀ ԳԱԱ ԻԱՊԻ ընդհանուր նիստ (2019թ.):

Հրապարակումներ: Ատենախոսության հիմնական արդյունքները հրատարակված են 7 գիտական աշխատություններում, որոնք թվարկված են սեղմագրի վերջում:

Աշխատանքի կառուցվածքը և ծավալը: Ատենախոսությունը բաղկացած է առաջաբանից, չորս գլխից, եզրակացությունից, 79 անուն օգտագործված գրականության ցանկից, թարգմանված տերմիններից և 8 էջ հավելվածից: Աշխատանքի ընդհանուր ծավալն է 122 էջ:

ԱՇԽԱՏԱՆՔԻ ԲՈՎԱՆԴԱԿՈՒԹՅՈՒՆԸ

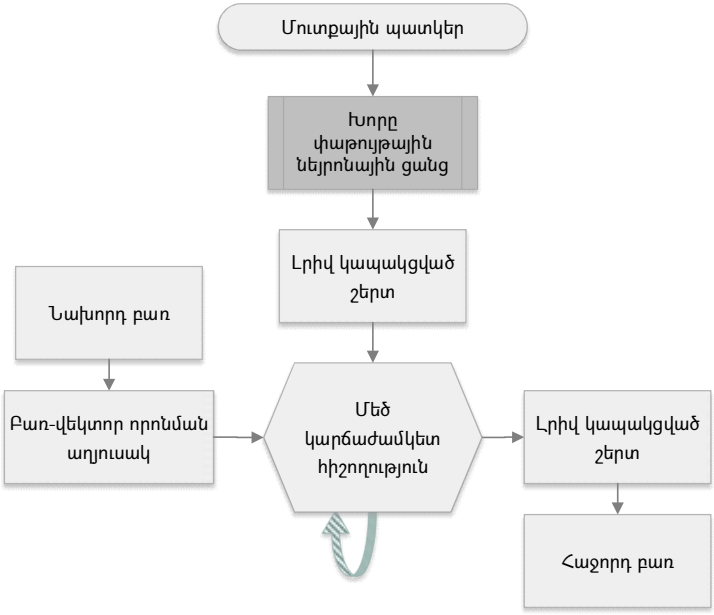
Առաջաբանում հիմնավորված է թեմայի արդիականությունը, ձևակերպված են աշխատանքի նպատակները, հետազոտության մեթոդները, գիտական և կիրառական նորույթները, որոնք ներկայացվում են պաշտպանության:

Առաջին գլխում ամփոփված են խորը փաթույթային նեյրոնային ցանցերով (deep convolutional neural networks) պատկերների իմաստաբանական վերլուծության մեթոդները և դրանց տարրերը:

Գլխում ամփոփված են ցանցերը կազմող ուսուցանվող (trainable) և չուսուցանվող շերտերը, մասնավորապես՝ փաթույթային (convolutional), միավորող (pooling), խրձով նորմալիզացնող (batch normalization) և լրիվ կապակցված (fully connected): Յուրաքանչյուր շերտի համար տրված է շերտի ելքային արժեքի հաշվարկման ֆունկցիան, ինչպես նաև նկարագրված է շերտի դերն ու կիրառությունները: Ամփոփված են նաև պատկերների իմաստաբանական վերլուծության խնդիրներում ցանցերի վերահսկվող ուսուցման եղանակը (supervised learning), երբ ցանցի ուսուցման (training) և վավերացման (validation) համար տրված են համապատասխան տվյալների հավաքածուներ (dataset), որոնք բաղկացած են պիտակավորված (labeled) օրինակներից: Անդրադարձ է արված բուն ուսուցման ժամանակ կիրառվող գրադիենտի անկման (gradient descent) և ետ տարածման (back propogation, տեղակայված է Հավելված 1-ում) մեթոդներին:

Երկրորդ գլխում ներկայացված են պատկերների վերնագրերի (caption) գեներացման խնդիրներում կիրառվող խորը փաթույթային նեյրոնային ցանցերի հատկությունները, ինչպես նաև ներկայացված է պատկերների վերնագրերի գեներատորի ընդհանուր կառուցվածքը, խնդրի մաթեմատիկական դրվածքը և ուսուցման ալգորիթմը: Ցանցերի հատկությունների հետազոտության շրջանակներում մշակվել է իմաստբանորեն նման պատկերների որոնման մեթոդ:

© National Library of Armenia



Պատկեր 1: Պատկերից վերնագիր գեներացնող նեյրոնային ցանցի մոդելի ընդհանուր սխեման:

Պատկերի վերնագրի գեներացումը կատարվում է հետևյալ հերթականությամբ (տես՝ Պատկեր 1).

- Խորը փաթույթային նեյրոնային ցանցը գեներացնում է իր մուտքին տրված պատկերը նկարագրող բարձր մակարդակի հատկությունները:
- Պատկերը նկարագրող բարձր մակարդակի հատկությունները տրվում են լրիվ կապակցված շերտի մուտքին: Շերտը տեղափոխում է այդ հատկությունները պատկեր-նախադասություն ներդրվածության տարածություն (մանրամասն՝ 2.3.2 կետում): Պատկեր-նախադասություն ներդրվածության տարածություն տեղափոխված հատկությունները մեծ կարճաժամկետ հիշողության բջջի (Long Short-Term Memory, LSTM) համար հանդիսանում են սկզբնական մուտքային վեկտոր:
- Կատարվում է սկզբնական մուտքային վեկտորի մշակում մեծ կարճաժամկետ հիշողության բջջի կողմից, որի արդյունքում ձևավորվում է դրա սկզբնական վիճակը:
- Այնուհետև կատարվում է պատկերի վերնագրի յուրաքանչյուր հաջորդ բառի գեներացիա:
  - Յուրաքանչյուր նախորդող բառ բերվում է պատկեր-նախադասություն ներդրվածության տարածություն: Արդյունքում ստացվում են մեծ կարճաժամկետ հիշողության բջջի մուտքերը:

© National Library of Armenia

- Յուրաքանչյուր իտերացիայում մեծ կարճաժամկետ հիշողության բջիջն իր մուտքին ստանալով մուտքային ազդանշանը, ձևավորում է ելքային վեկտորը և վիճակը:
- Յուրաքանչյուր ելքային վեկտորից հաշվարկվում է նախադասության ընթացիկ բառի հավանականությունների բաշխումը բառարանի բոլոր բառերի վրա: Բաշխումից կախված ընտրվում է նախադասության հաջորդ բառը:

Իմաստաբանորեն նման պատկերների որոնում: Խորը փաթույթային նեյրոնային ցանցերի միջոցով գեներացված պատկերների բարձր մակարդակի հատկությունների հետազոտման նպատակով մշակվել է նման պատկերների որոնման մեթոդ, որը ներկայացված է 2.2.2 կետում: Յուրաքանչյուր փնտրվող պատկերի համար հաշվարկվում են բարձր կարգի հատկություններ և համեմատվում բազայում գտնվող յուրաքանչյուր պատկերի բարձր կարգի հատկությունների հետ, այնուհետև որպես նման պատկերներ դիտարկվում են ամենամոտ գտնվողները: Դիցուք, GoogLeNet ցանցում բարձր մակարդակի հատկություններն իրենցից ներկայացնում են պատկերի պարունակության՝ 1024 երկարությամբ իրական թվերի վեկտոր: Այդ վեկտորն այնքան ինֆորմացիա է պարունակում պատկերի մասին, որ հնարավոր է կատարել նման պարունակությամբ պատկերների որոնում 89% ճշգրտությամբ, երբ հեռավորության մետրիկան կորելացիան է: Պատկեր 2–ում ներկայացված են որոնման արդյունքները, որոնք ունեն տեսողական նմանություններ՝ պարունակում են միանման օբյեկտներ՝ օրինակ լուսին, երկինք, ամպեր, դաշտ: Նշենք, որ պատկերում առարկաների մի մասը բացակայում է ուսուցման տվյալների հավաքածուի պիտակների բազմությունից:

Պատկեր 2: Ձախում՝ որոնվող պատկերին հաջորդում են որոնման արդյունքները:

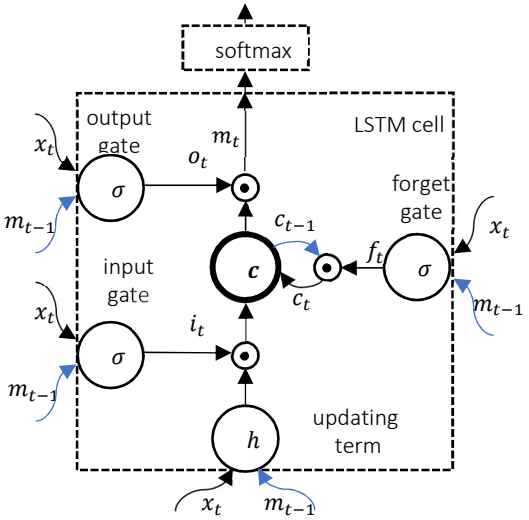
Մեծ կարճաժամկետ հիշողության բջջի կորիզը (տես՝ Պատկեր 3) հանդիսանում է *c* հիշողությունը, որը կոդավորում և պահում է մինչ այդ քայլը ստացած ինֆորմացիան: Բջջի վարքագիծը ղեկավարվում է «դարպաս (gate)»-ներով, որոնք կախված են նախորդ ելքից և ընթացիկ մուտքից: Դարպասները ընդունում են 0–ից 1 արժեքներ, որի շնորհիվ մուտքային (inpout gate) և ելքային (output gate) դարպասները կարողանում են կառավարել մուտքային և ելքային արժեքները: Մոռացության դարպասը (forget gate) որոշում է, թե

© National Library of Armenia

բջջի նախորդ արժեքի որ մասը պետք է պահպանել: Ֆորմալ նկարագրությունը տրվում է հետևյալ հավասարումներով՝

$$
\begin{aligned}
i_t &= \sigma(W_{ix}x_t + W_{im}m_{t-1}),\\
f_t &= \sigma\left(W_{fx}x_t + W_{fm}m_{t-1}\right),\\
o_t &= \sigma(W_{ox}x_t + W_{om}m_{t-1}),\\
c_t &= f_t \odot c_{t-1} + \ i_t \odot h(W_{cx}x_t + W_{cm}m_{t-1}),\\
m_t &= o_t \odot c_t,
\end{aligned}
$$

որտեղ i_t-ն ու o_t–ն մուտքային և ելքային դարպասներն են, f_t-ն մոռացության դարպասն է, $\odot$-ը դարպասի հետ արտադրյալներն են, c_t–ն՝ t պահին բջջի վիճակն է, m_t-ն՝ ռեկուրենտ ցանցի t-րդ իտերացիայի ելքը, W_{ix}-ը, W_{im}-ը, W_{fx}-ը, W_{fm}-ը, W_{ox}-ը, W_{om}-ը, W_{cx}-ը, W_{cm}-ը, W_{pm}-ը՝ մեծ կարճաժամկետ հիշողության բջջի ուսուցանվող կշիռներն են, $\sigma(\cdot)$ և $h(\cdot)$ համապատասխանաբար սիգմոիդ և հիպերբոլիկ տանգենս ոչ գծային ֆունկցիաները:



Պատկեր 3: LSTM - Մեծ կարճաժամկետ հիշողության բջջի գծապատկերը:

Երրորդ գլխում ներկայացված են պատկերների վերնագրերի գեներատորի ուսուցման և գոծարկման համար մշակված ծրագրային համակարգերը, ռեկուրենտ նեյրոնային ցանցի ուսուցման քայլերը, փորձերը և ստացված արդյունքները:
Պատկերների վերնագրեր գեներացնող մոդելի իրականացման համար կատրավել է.

- Փաթույթային և ռեկուրենտ նեյրոնային ցանցերի միասնական մոդելի ծրագրային իրականացում,
- Ցանցին ուսուցող մոդելի ծրագրային իրականացում,

© National Library of Armenia

- Ուսուցման և վավերացման համար նախատեսված տվյալների հավաքածուների մշակում,
- Ուսուցող և վավերացնող ծրագրերի իրականացում,
- Պատկերների վերնագրեր գեներացնող ցանցի ուսուցում,
- Նեյրոնային ցանցի գործարկման համար ծրագրային համակարգի մշակում:

Նեյրոնային ցանցի մոդելի արագ ուսուցանման նպատակով, ստեղծվել է ծրագրային համակարգ, որը կարողանում է սպառել մեծ հաշվողական հզորություն: Աշխատանքի շրջանակներում օգտագործվել է TensorFlow բաց ծրագրային կոդով գրադարանը, որը հնարավորություն է տալիս կատարել թվաբանական հաշվարկներ՝ կիրառելով տվյալների հոսքերի գրաֆներ (data flow graphs): Պատկերների հետ աշխատանքը կազմակերպելու նպատակով կիրառվել են numpy, OpenCV, matplotlib գրադարանները: Պատկերների վերնագրեր գեներացնող համակարգի ստեղծման համար կատարվել է մոդելի ուսուցանում: Ցանցերի ուսուցման նպատակով օգտագործվել է MSCOCO պատկերների բազան, որը պարունակում է ուսուցման և վավերացման համար նախատեսված նախորոք վերնագրված մեծ թվով պատկերներ: MSCOCO պատկերների բազան բաց է և կարելի է ներբեռնել համացանցից: Ուսուցման և վավերացման պատկերների հավաքածուների պիտակները նշված են մարդկանց կողմից: Մասնավորապես այդ հավաքածուներում յուրաքանչյուր պատկերին համապատասխանեցված է 4-8 վերնագիր:

3.2 ենթագլխում ներկայացված է MSCOCO պատկերների ուսուցման և վավերացման հավաքածուների նախնական մշակման ալգորիթմը, որի արդյունքում ստացվում են ուսուցման և վավերացման օրինակների խրձերի հոսքերը (stream), որոնք մուտք են հանդիսանում համապատասխանաբար ուսուցման և վավերացման ծրագրային համակարգերին:

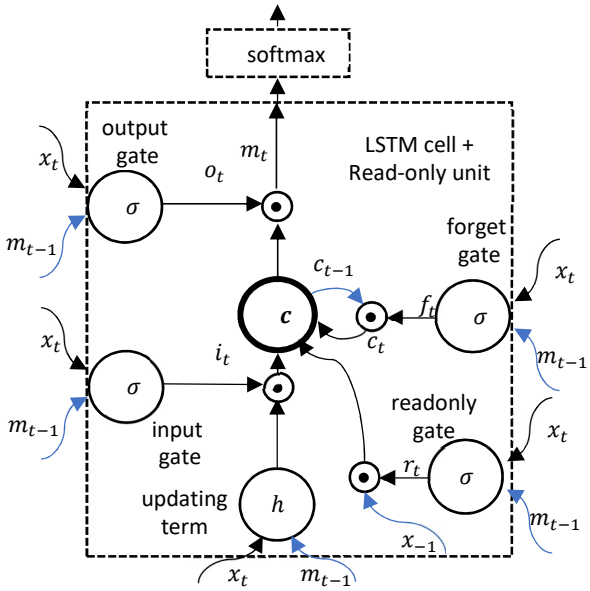
3.3 ենթագլխում ներկայացված են պատկերների վերնագրեր գեներացնող մոդելի ուսուցման և վավերացման համակարգերի ծրագրային իրականացումների նկարագրությունները:

3.4 և 3.5 ենթագլխներում ներկայացված են պատկերների վերնագրեր գեներացնող մոդելի ուսուցման, վավերացման գործընթանցները և ստացված արդյունքները: Ինչպես նաև ներկայացված են մշակված ծրագրային համակարգեր, որոնց միջոցով կարելի է կիրառել ռեկուրենտ նեյրոնային ցանցերը և՛ պատկերների, և՛ տեսաշարերի, և՛ տեսախցիկների համար:

3.6 ենթագլխում ներկայացված է ռեկուրենտ ցանցերի ուսումնասիրության համար մշակված ծրագրային համակարգ, որը թույլ է տալիս ավելի քան 10 անգամ կրճատել ուսուցման մեկ փորձի վրա ծախսվող ժամանակը՝ միևնույն ժամանակ նվազեցնելով գրաֆիկական քարտի հիշողության օգտագործումը: Մեծ կարճաժամկետ հիշողության հնարավորությունների մեծացման նպատակով, ստեղծվել է հետևյալ մոդիֆիկացիան: Ավելացվել է ևս մեկ դարպաս (տես՝ Պատկեր 4), որը մասնակցում է նոր վիճակի արժեքի որոշմանը: Դարպասը կանվանենք՝ ընթերցման (readonly gate), իսկ բջիջը՝ հաստատուն բաղադրիչով: Դարպասը հաշվարկվում է ըստ (1) հավասարման, որտեղ W_{rx} և W_{rm}-ն

© National Library of Armenia

ուսուցանվող կշիռներ են: Դարպասը մասնակցում է բջջի վիճակի հաշվարկին՝ (2) հավասարմամբ կարգավորելով սկզբնական մուտքային վեկտորի չափաբաժինը:



Պատկեր 4: Հաստատուն բաղադրիչով մեծ կարճաժամկետ հիշողության բջջի գծապատկերը:

$$r_t = \sigma(W_{rx}x_t + W_{rm}m_{t-1}) \quad (1)$$

$$c_t = f_t \odot c_{t-1} + r_t \odot x_{-1} + i_t \odot h(W_{cx}x_t + W_{cm}m_{t-1}) \quad (2)$$

Այսպիսով, մշակված հաստատուն բաղադրիչով մեծ կարճաժամկետ հիշողության բջիջ ունեցող ռեկուրենտ նեյրոնային ցանցը ուսուցանվել է MSCOCO պատկերների ուսուցման տվյալների հավաքածուի վրա և ունեցել է 1.85 կորստի ֆունկցիայի արժեք, այն դեպքում, երբ չմոդիֆիկացված մեծ կարճաժամկետ հիշողության համար այն հավասար է եղել 2.15: Իսկ վավերացման տվյալների հավաքածուի համար դրանք հավասար են եղել 1.90 և 2.05՝ համապատասխանաբար: Թեստավորման ժամանակ ստացվել է, որ հաստատուն բաղադրիչով մեծ կարճաժամկետ հիշողությունը որոշ դեպքերում ավելի լավ է գեներացնում պատկերների վերնագրերը (տես՝ Պատկեր 5):

© National Library of Armenia

a woman sitting at a table with a laptop.
a woman sitting at a table with a laptop
a woman sitting at a table in front of a laptop .

a woman sitting at a table in a dress
a woman is sitting at a table with a hat.
a woman sitting at a table with a hat on .

a woman in a bikini standing on a beach.
a woman in a bikini standing on a beach holding an umbrella.
a young girl in a bikini standing on a beach.

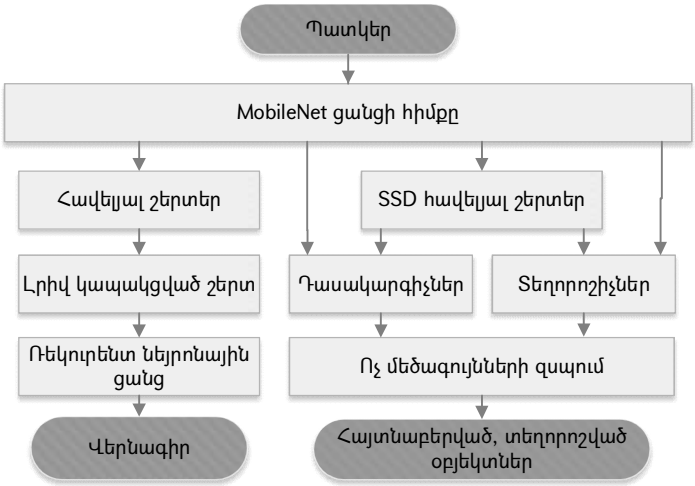
a woman in a dress and a hat on a beach.
a girl in a dress and a hat on a beach
a woman in a dress and a hat on a beach

Պատկեր 5: Երկու տարբեր մոդելների կողմից գեներացված պատկերների վերնագրերի համեմատությունը: Կապույտ վերնագրերը համապատասխանում են մեծ կարճաժամկետ հիշողությանը, կարմիրները՝ հաստատուն բաղադրիչով մեծ կարճաժամկետ հիշողությանը:

Չորրորդ գլխում ներկայացված է պատկերների վերնագրեր գեներացնող ռեկուրենտ նեյրոնային ցանց, որը հիմնված է օբյեկտներ հայտնաբերող և տեղորոշող խորը փաթույթային նեյրոնային ցանցից դուրս բերված բարձր մակարդակի հատկությունների վրա, ինպես նաև բարձր արագագործության ծրագրային համակարգ, որը կատարում է պատկերների վերնագրերի գեներացում, օբյեկտների հայտնաբերում, ճանաչում և տեղորոշում, ինչպես նաև հաշվում է օբյեկտների գերիշխող գույները:

Մշակվել է մեկ միասնական ցանց (տես՝ Պատկեր 6), որը մուտքային պատկերների համար գեներացում է վերնագրերը, հայտնաբերում, ճանաչում և տեղորոշում է օբյեկտները:

© National Library of Armenia



Պատկեր 6: Օբյեկտների հայտնաբերման և տեղորոշման MobleNet-SSD ցանցի վրա հիմնված պատկերների վերնագրերի գեներատորի սխեման:

Ցանցը կազմված է օբյեկտների հայտնաբերման և տեղորոշման նախորոք ուսուցված MobileNet-SSD ցանցից, ուսուցվող հավելյալ շերտերից (տես՝ Աղյուսակ 1) և ռեկուրենտ նեյրոնային ցանցից: Ցանցը ուսուցանվել է MSCOCO պատկերնեի հավաքածուի վրա:

Աղյուսակ 1: Օբյեկտների հայտնաբերման և տեղորոշման MobleNet-SSD ցանցի վրա հիմնված պատկերների վերնագրերի գեներատորի սխեմայում ավելացված հավելյալ շերտերը: (Conv – 2D Convolution, Conv dw – Depthwise separable convolution)

Տեսակ	*Ցատկ*	*Ֆիլտրի չափ*	*Մուտքային չափ*
Conv dw	2	3 x 3 x 512 dw	19 x 19 x 512
Conv	1	1 x 1 x 512 x 1024	10 x 10 x 512
Conv dw	1	3 x 3 x 1024 dw	10 x 10 x 1024
Conv	1	1 x 1 x 1024 x 1024	10 x 10 x 1024

Ստացված ցանցի համար հաշվարկվել են BLEU 1, BLEU 2, BLEU 3, BLEU 4, METEOR, ROUGE-L, CIDE մետրիկաները և կատարվել է համեմատություն լավագույն արդյունքի հետ (տես՝ Աղյուսակ 2): Համեմատության մեջ նաև ներառված են մարդկանց կողմից առաջարկված վերնագրերի մետրիկաները: Հաշվարկվել են բառարանից պատահական բառերով կազմված վերնագրերի (ճիշտ երկարությամբ) և ճիշտ վերնագրերի բառերը պատահականորեն վերադասավորած նախադասությունների մետրիկաները: Վերջին երկուսը հաշվարկվել են, որպեսզի հնարավոր լինի հեշտությամբ հասկանալ, թե յուրաքանչյուր մետրիկան ինչպես է գնահատում պատահականորեն գեներացված նախադասությունները:

© National Library of Armenia

Աղյուսակ 2: Պատկերներում օբյեկտների հայտնաբերման և տեղորոշման MobileNet-SSD ցանցի վրա հիմնված պատկերների վերնագրերի գեներատորի ճշգրտության մետրիկաների համեմատության աղյուսակ:

	BLEU 1	BLEU 2	BLEU 3	BLEU 4	METEOR	ROUGE-L	CIDE
MobileNet հիմքով	0.491	0.296	0.181	0.115	0.151	0.379	0.256
Vinyals	-	-	-	0.277	0.237	-	0.855
Մարդ	0.663	0.469	0.321	0.217	0.262	0.484	0.854
Պատահական	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Խառնած	1.0	0.404	0.139	0.050	0.415	0.469	1.102

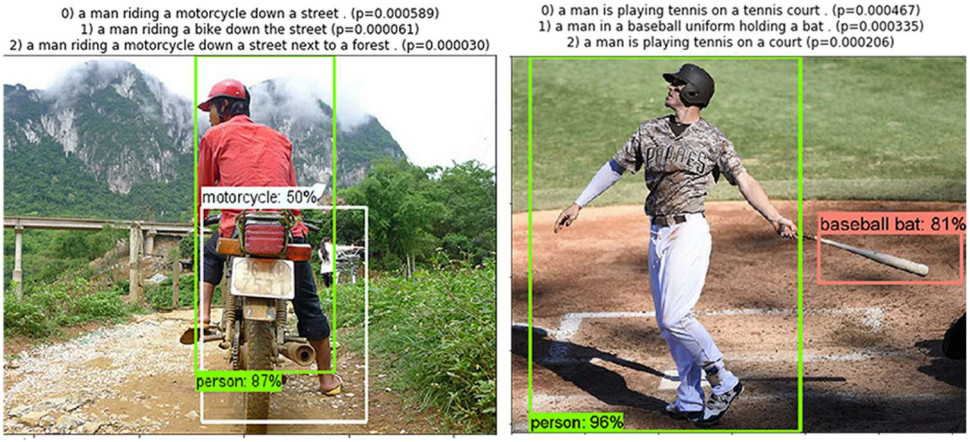
Աղյուսակ 2-ում ներկայացված մետրիկաներից երևում է, որ ստացված ցանցը ճշգրտությամբ զիջում է GoogLeNet ցանցի վրա կառուցված վերնագրերի գեներատորին: Ստեղծված ցանցը գործարկելու նպատակով՝ մշակվել է ծրագրային համակարգ, որը մուտքին ստանում է պատկերները (տեսաշարը կամ տեսախցիկի իդենտիֆիկացիոն համարը) և գեներացնում է այդ պատկերները նկարագրող վերնագրեր, հայտնաբերում է պատկերներում գտնվող օբյեկտները և տեղորոշում դրանք:

Փորձարկելով մշակված ցանցը և ծրագրային համակարգը՝ ստացվել են հետևյալ արդյունքները: Համակարգը որոշ պատկերների համար կարողանում է բավականին լավ գեներացնել վերնագրեր (տես՝ Պատկեր 7): Որոշ պատկերների համար երևում է, որ ցանցը վերնագրերը գեներացնելիս՝ «կին» բառի փոխարեն կիրառում է «տղամարդ» բառն (տես՝ Պատկեր 8) այն պարզ պատճառով, որ MobileNet-SSD ցանցում առկա է միայն «անձ» հասկացությունը: Ցանցի նման վարքագիծը կարելի է բացատրել ելնելով այն հանգամանքից, որ MobileNet-SSD օբյեկտների հայտնաբերող և տեղորոշող ցանցը ուսուցանված է 100 դասի օբյեկտների համար, իսկ պատկերների վերնագրերը պարունակում են ավելի քան 11000 բառ Այսպիսով ստացվում է, որ մեծ կարճաժամկետ հիշողությունը փորձում է ապակոդավորել 11000 բառերը՝ օգտվելով ընդամենը 100 բառ նկարագրող հատկությունների վեկտորից (Աղյուսակ 1-ի վերջին շերտը): Հենց այդպիսի տարբերությունն էլ Աղյուսակ 2-ում ներկայացված ցածր ճշգրտության պատճառ է հանդիսանում: Այնուամենայնիվ, MobileNet-SSD ցանցում կոդավորվող հասկացությունների շրջանակում, վերնագրերի գեներատորը լուծում է իր առջև դրված խնդիրը:

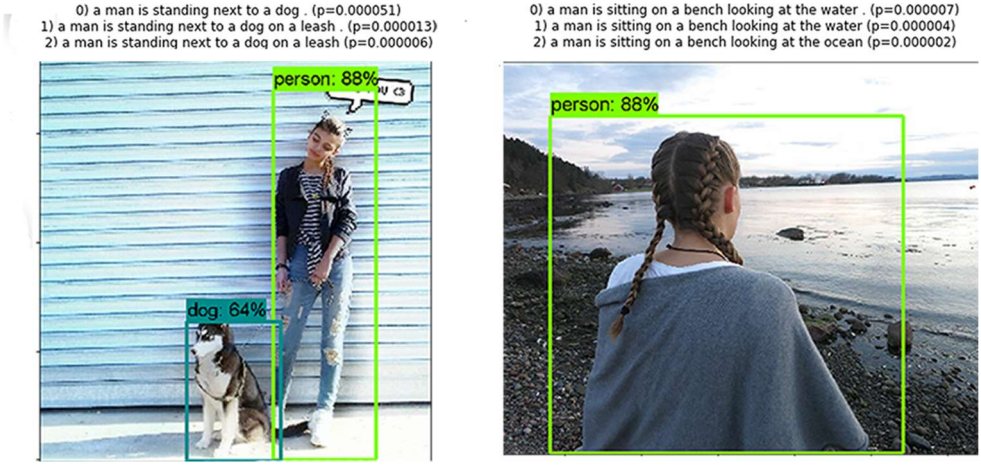
Նշենք, որ համակարգի ճշգրտության մեծացման նպատակով կարելի է վերնագրեր գեներացնող ցանցը ուսուցանել ավելի մեծ քանակի օբյեկտների դասեր հայտնաբերող և տեղորոշող ցանցերի վրա: Նշենք նաև, որ ցանցի մեծացման հետ մեկտեղ մեծանում է ուսուցմանն անհրաժեշտ հիշողությունը, որը ապահովելու համար անհրաժեշտ են ավելի հզոր տեխնիկական լուծումներ, քան աշխատանքում օգտագործված մեկ GTX 1080 գրաֆիկական քարտը:

Արդյունքում ստացվել է միասնական ցանց (մոդել), որի միջոցով Intel Core i5-9600K պրոցեսորի վրա մեկ պատկերի մշակումը միջինում տևում է 0.67 վայրկյան, այնինչ, անջատ ցանցերի զուգահեռ աշխատանքը տևում է 0.74 վայրկյան:

© National Library of Armenia



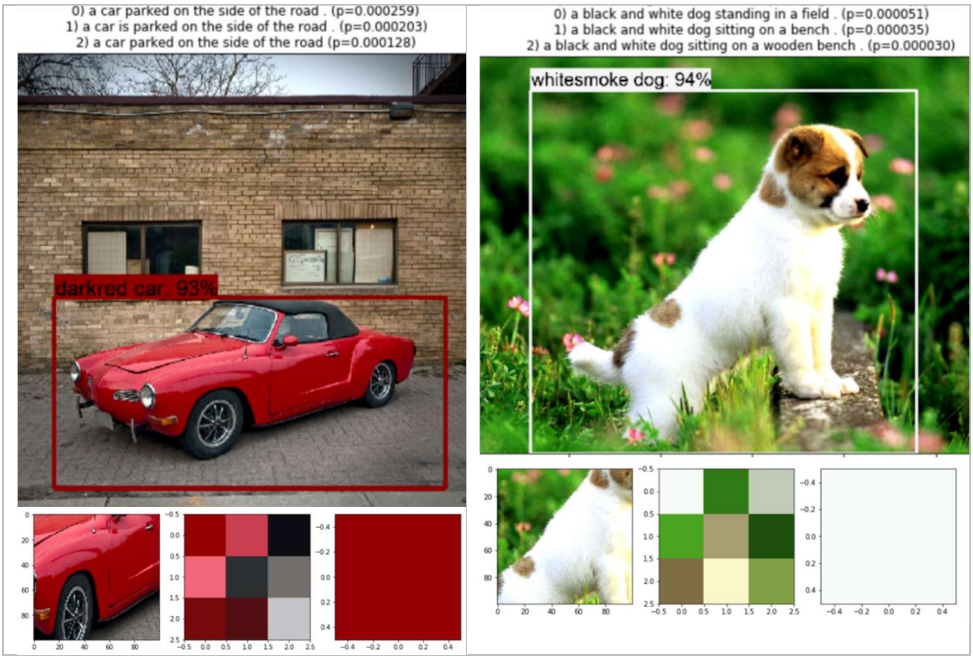
Պատկեր 7: Պատկերների վերնագրերի գեներացում, օբյեկտների հայտնաբերում և տեղորոշում: Ներկայացված են ճիշտ վերնագրերի գեներացման դեպքեր:



Պատկեր 8: Պատկերների վերնագրերի գեներացում, օբյեկտների հայտնաբերում և տեղորոշում: Ներկայացված են վերնագրերի թերի գեներացման դեպքեր:

Այսպիսով ստեղծվել է պատկերների վերնագրեր գեներացնող, պատկերներում օբյեկտները հայտնաբերող և տեղորոշող մեկ միասնական արագագործ ցանց: Մշակվել է համապատասխան ծրագրային համակարգ, որի կիրառմամբ հնարավոր է իրական ժամանակում գեներացնել պատկերների վերնագրերը, հայտնաբերել պատկերում օբյեկտները և տեղորոշել դրանք, ինչպես նաև հայտնաբերված օբյեկտներին տալ գունային նկարագրիչներ (տես՝ Պատկեր 9):

© National Library of Armenia



Պատկեր 9: Պատկերների վերնագրերի գեներացում, պատկերում օբյեկտների հայտնաբերում և տեղորոշում, օբյեկտների գույնի հայտնաբերում: Երկրորդ տողում ներկայացված են օբյեկտների գերիշխող գույնի հայտնաբերման քայլերը:

© National Library of Armenia

ԱՇԽԱՏԱՆՔԻ ՀԻՄՆԱԿԱՆ ԱՐԴՅՈՒՆՔՆԵՐԸ

- Մշակվել է պատկերների վերնագրեր գեներացնող ռեկուրենտ նեյրոնային ցանց, որը հիմնված է օբյեկտներ հայտնաբերող և տեղորոշող խորը փաթույթային նեյրոնային ցանցից դուրս բերված բարձր մակարդակի հատկությունների վրա: Արդյունքում ստեղծվել է պատկերների վերնագրեր գեներացնող, պատկերներում օբյեկտներ հայտնաբերող և տեղորոշող մեկ միասնական, արագագործ ցանց: Ցանցերի ուսուցման, գնահատման և թեստավորման համար ստեղծվել են ծրագրային համակարգեր, որոնք հաշվարկները զուգահեռ կատարում են կենտրոնական և գրաֆիկական պրոցեսորների բոլոր միջուկների վրա: Մշակվել է իրական ժամանակում աշխատող ծրագրային համակարգ, որը կատարում է պատկերների վերնագրերի գեներացում, օբյեկտների հայտնաբերում, ճանաչում և տեղորոշում, ինչպես նաև հաշվում է օբյեկտների գերիշխող գույները: [1-3]

- Մշակվել է իմաստաբանորեն նման պատկերների որոնման ալգորիթմ, որը GoogLeNet ցանցից դուրս բերված բարձր մակարդակի հատկությունների կիրառմամբ կատարում է օբյեկտների որոնում 89% ճշգրտությամբ: [4]

- Մշակվել է մեծ կարճաժամկետ հիշողության (Long Short-Term Memory) մոդիֆիկացված տարբերակ, որը պատկերների վերնագրերի գեներացման խնդրում ցուցաբերել է ավելի բարձր ճշգրտություն: Ուսումնասիրության համար ստեղծվել է ծրագրային համակարգ, որը թույլ է տալիս մոտ 10 անգամ կրճատել ուսուցման մեկ փորձի վրա ծախսվող ժամանակը և նվազեցնել գրաֆիկական քարտի հիշողության օգտագործումը: [5-7]

© National Library of Armenia

ՀՐԱՏԱՐԱԿՎԱԾ ԱՇԽԱՏՈՒԹՅՈՒՆՆԵՐԸ

[1] *A. Poghosyan, H. Sarukhanyan, " Image Caption Generation model based on Object Detector", Mathematical Problems of Computer Science vol. 50, pp. 5–14, 2018.*

[2] *A. Poghosyan, "Image Caption Generation and Object Detection via a Single Model", Mathematical Problems of Computer Science vol. 48, pp. 42–49, 2017.*

[3] *N. Hayrapetyan, R. Hakobyan, A. Poghosyan, V. Gabrielyan, "Border Surveillance Using UAVs with Thermal Camera", NATO Science for Peace and Security Series – D: Information and Communication Security, vol. 47, pp. 219 – 225, 2016.*

[4] *A. Poghosyan, H. Sarukhanyan, "Image Visual Similarity Based on High Level Features of Convolutional Neural Networks", Mathematical Problems of Computer Science, vol. 45, pp. 138-142, 2016.*

[5] *A. Poghosyan, H. Sarukhanyan, "Long Short-Term Memory with Read-only Unit in Neural Image Caption Generator", 11-th International Conference Computer Science and Information Technologies, (CSIT-11), 2017, pp. 368-371.*

[6] *A. Poghosyan, H. Sarukhanyan, "Short-term memory with read-only unit in neural image caption generator", 11-th International Conference Computer Science and Information Technologies, Revised Selected Papers, IEEE Xplore, 10.1109/CSITechnol.2017.8312163, Electronic ISBN: 978-1-5386-2830-0, Print on Demand(PoD) ISBN: 978-1-5386-2831-7, pp. 162- 167, 2017*

[7] *A. Poghosyan, H. Sarukhanyan, "RNN with Additional Constant Memory for Image Caption Generation Task", International Academy Journal Web of Scholar, 4(13), Vol.1, pp. 3–7, 2017.*

© National Library of Armenia

Poghosyan Aghasi Stepan

DEVELOPMENT AND IMPLEMENTATION OF METHODS OF DIGITAL IMAGE SEMANTIC ANALYSIS

RESUME

The relevance of the work. The term “semantic image analysis” means the derivation of a meaningful, figurative description of the content of an image, similar to what a person could give. The implementation of technology implies the creation of a system which will produce results in the form of elements of a natural language - words and sentences for incoming images. At the same time, the individual pixel values of the image, for example, the color of a pixel or the brightness of an image, cannot be considered as its semantic description, but can be a part of a phrase or a sentence, for example, “red car” or “a man is sitting in a bright room”.

Today, this technique is used in development process of search and recommendation systems, self-driving vehicles and unmanned aerial vehicles. Also, it can be used for developing tools which could help people with vision problems.

The solution of the problem comes into contact with such areas as machine learning, statistics, natural language processing, computer vision, signal and image processing, as well as information theory and neurobiology.

The complexity of the problem lies in the fact that image processing and pattern recognition methods alone are not enough to perform semantic image analysis, in particular object detection and localization since the objects in the image could be considered as different objects depending on their location. For example, hair of an animal may be considered as a web.

Various systems of semantic image analysis have been created during recent years. The state-of-the-art approaches are based on deep convolutional neural networks. There are networks that can both detect and localize the objects in the image, as well as generate captions in the form of a natural language.

The aim of this work is to develop a real-time method, which will caption the image via sentence expressed in a natural language such as English, also will detect and localize objects and compute their color attributes. Furthermore, the goal of the work is to develop a software system, to accelerate the calculations by utilizing all cores of the CPU and the GPU.

The main results.

- A caption generator and object detector have been developed as a single model by mixing a recurrent neural network and object detector. The result is a single high-speed network that can generate image captions, detect and localize objects. Software systems have been developed for training, evaluating and testing the recurrent neural network. A real-time software has been developed that generates image caption, detects objects and localizes them, also calculates the dominant color. The software performs computations in parallel by utilizing all the cores of the CPU and the GPU. [1-3]

© National Library of Armenia

- A method was developed for finding similar images, which has 89% accuracy when GoogLeNet's high-level feature's metric is a correlation. [4]
- A modified version of the Long Short-term Memory has been developed, which shows better results in image caption generation task. A software system was developed that allows reducing the training duration of the recurrent network by more than 10 times and also the usage of a GPU memory. [5-7]

© National Library of Armenia

Погосян Агаси Степанович

РАЗРАБОТКА И РЕАЛИЗАЦИЯ МЕТОДОВ СЕМАНТИЧЕСКОГО АНАЛИЗА ЦИФРОВОГО ИЗОБРАЖЕНИЯ

РЕЗЮМЕ

Актуальность работы. Термин «семантический анализ изображений» означает вывод осмысленного, образного описания содержания изображения – такого, какое бы смог дать человек. Реализация техники подразумевает создание системы, которая для входящих изображения будет выдавать результаты в виде элементов естественного языка – слов и предложений. При этом отдельно взятые физические величины изображения, например, цвет пикселя или яркость изображения не могут считаться его семантической оценкой, но могут составлять части словосочетания или предложения, например, «красная машина», «мужчина сидит в яркой комнате».

На сегодняшний день эта техника используется в таких областях, как разработка поисковых и рекомендательных систем, автопилотируемых автомобилей и беспилотных летательных аппаратов. Также, при правильном применении, с ее помощью можно разработать средства, помогающие людям с проблемным зрением.

Решение проблемы соприкасается с такими областями, как машинное обучение, статистика, обработка естественного языка, компьютерное зрение, обработка сигналов и изображений, а также теория информации и нейробиология.

Сложность проблемы состоит в том, что выявление и позиционирование объектов, которое составляет часть смыслового анализа изображений, не решается только методами обработки изображений и распознавания образов, поскольку объекты в изображении, в зависимости от расположения относительно друг друга, могут рассматриваться как разные предметы. Например, отдельно взятые волоски животного могут быть похожи на паутину.

В последние годы были созданы различные системы смыслового анализа изображений. Из числа которых системы, основанные на сверточных нейронных сетях. На их основе разработаны сети, способные как выявлять, распознавать и обнаруживать объекты в изображениях, так и генерировать заглавия изображений в виде естественного языка.

Целью работы является разработка быстродействующего метода, с помощью которого возможно озаглавить изображение предложением, соответствующим грамматике естественного языка, распознать и обнаружить объекты в изображении, а также рассчитать преобладающий цвет для них. Также целью работы является создание программного обеспечения, основанного на вышеописанном методе, которое, в целях ускорения расчетов, будет использовать все ядра как центрального, так и графического процессоров.

Основные результаты.

- Была разработана генерирующая заглавия изображений рекуррентная нейронная сеть, основанная на свойствах высокого уровня, выведенных из глубоких

© National Library of Armenia

сверточных нейронных сетей, распознающих и обнаруживающих объекты. В результате получилась единая быстродействующая сеть, способная генерировать заглавия изображений, распознавать и обнаруживать объекты в них. Были разработаны системы для оценки и тестирования обучения нейронных сетей, которые производят расчеты параллельно на всех ядрах центрального и графического процессоров. Было разработано программное обеспечение с высокой производительностью, которое выполняет генерацию заглавий изображений, выявляет, распознает и обнаруживает объекты, а также рассчитывает преобладающий в них цвет. [1-3]

- Был разработан алгоритм для нахождения изображений похожих по смыслу, который используя свойства высокого уровня, выведенные из сети GoogLeNet, реализует поиск объекта с точностью 89%. [4]
- Был разработан модифицированный вариант долгой краткосрочной памяти, который показывает лучшие результаты в генерации заглавий изображений. Была разработана система, позволяющая сокращать время необходимое для обучения примерно в 10 раз, используя меньше графической памяти. [5-7]

© National Library of Armenia