ARGUS Media Based Intelligence

PANOPTES – Automated Article Segmentation of Newspaper Pages for "Real Time Print Media Monitoring" M. Arnold, M. Cieliebak, T. Stadelmann, J. Stampfli, and F. Uzdilli





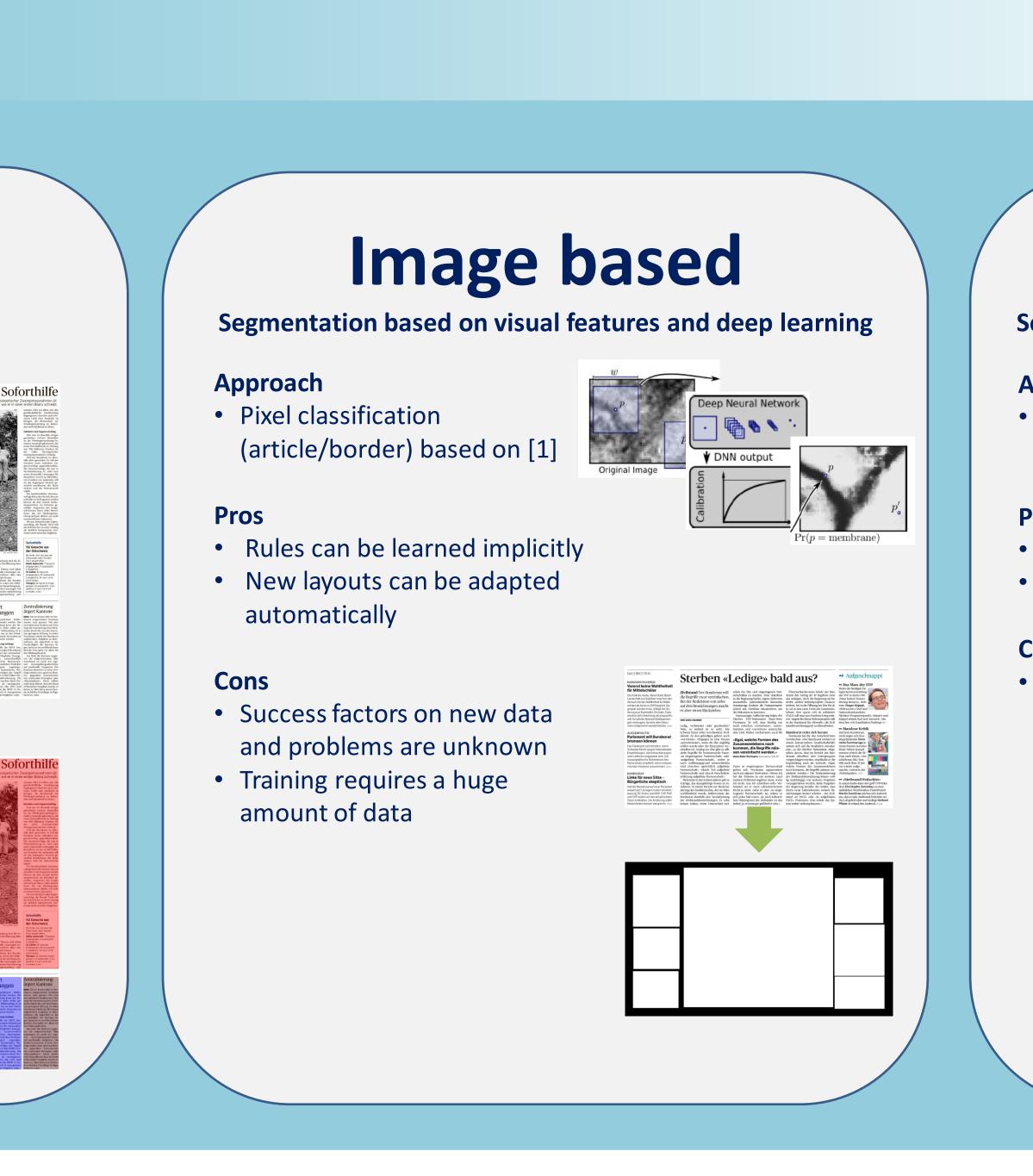
This project was funded by CTI under project number 17719.1 PFES-ES

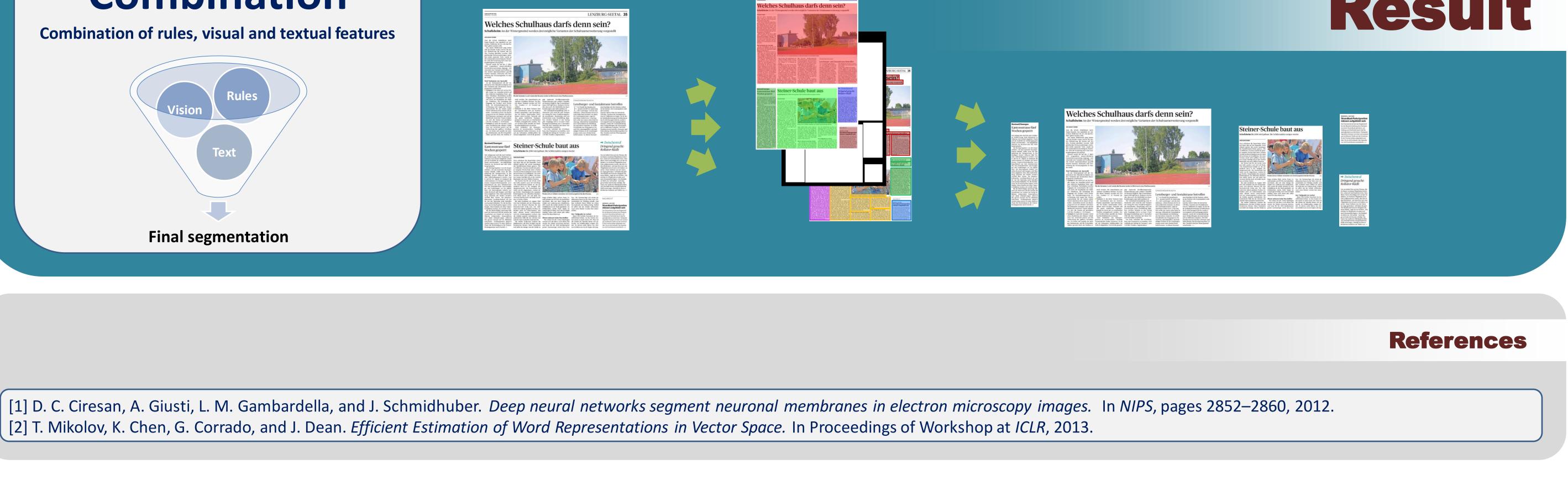
Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER Commission for Technology and Innovation CTI Innovation Promotion Agency

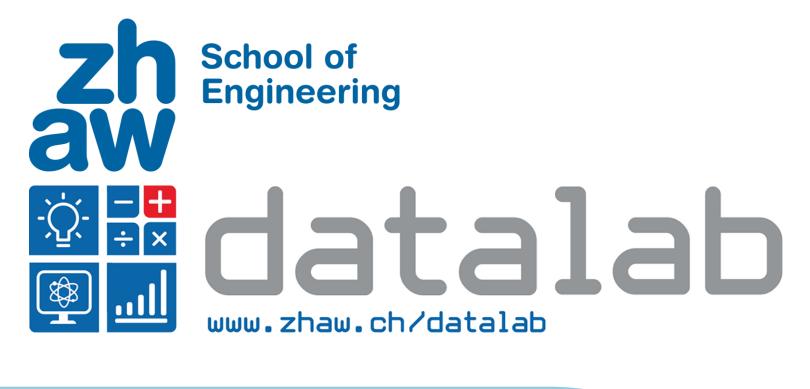
Overview **Rule based** Segmentation based on hardcoded rules **Rule examples** ther Polizei Über 1300 Gesuche um Soforthilfe • Each article must contain a title • Titles define article's width • Articles are graphically separated by e.g. lines • etc. Pros Performance increases the more time is spent for finding rules Adding new rules is simple Cons • Not every case can be covered Adaptation to new layouts is costly manual work EU unterstützt Freihandelsabkommen Strassburg, in einen zweiten Anlast het des Europaparionen gestem grundsätzlich für das Prahastala-alseennes mit der USA gestimmt. Bei deser erneuten Abstimmung wurde beschlossen, das unstittlume Schledb-gerichtevarlahren bahan zu lassen. Netiber wanne über des Erstehet nicht begesteht, die das Dornokastisptisch Serbischer Präsident bei Gedenkfeier dabei Beignant, Britterer Holdmann Beignant, Britterer Holdmann terpflicker Assamler Vuorenern an der Gederickersemmerins zum 2011. Jahres-tage des Massakes in Stebenick in der bosinscham Eindet sie. Flussen brachte eilert inge mit seinern Vuor eine Resulu-tion des UND-Schertrettalsteine zu Feit, mit der von-Völkermord- die Rede ist. Der Beit prangeweit aus tarkt die Weighten von berben im Beamlen-Krieg an. 504 Farc will einen Monat Waffenruhe Hexana, Die Unterhänder der kolum-bereichen Fan-Quarfin teiser tal Proderessenstendungen einen Mona Vielfenute angeköndigt. Die Waffen-nah solie ab dari 20. unligsten, segte der Fan-Chalansheiden bein Mörquez Jaumalaten. Man wohe der Fondeungen deren gest Ditaten Rech-nung teigen, ein den Friedemaprozes unterstützen. Die Fans kampt seit guf 90 Jahren gegen den Staat. sich Combination Bertusconi wegen Bestechung verurteilt Mappel, Der Hermalige traiteritache Ministerpoteistierten Silvio Berlussoni tit georgen in Naspel wegen der Bedu Hang einen Beraufenz auf auf Ahmen Hatt veranteilt wortfert. Dass der The-Jähnige winich in Sechligte muss, ist unwahrscheinket. Die Berlug ein Beraufen geiningen kann, wied de ter bei verpflicht zum, Dem Medonmilike dier weste vorgeworten, dem trähmen Infeltionen Euro, gezahlt zu hatten. Die **Combination of rules, visual and textual features** Südsudans Präsident wieder drei Jahre im Amt Judies, Gestern, kurz vor dem vierten Judirestig der Umbihörgigkeit wurde Bisfausdarn Präcernt Bien ein der Umbihörgigkeit Nicht masselle stähnsicht nachteiten er dem Partamern einderigen, die eine Anth Versteh ministen. Für die Zusantt wa-geneh er, dem Finderin zu anfahresten und ist dem Rebelens Recoten zu anfahresten um Fischellen Rebelen zu anfahresten. Son Balionsfigur der Tea Party Todesurteile nach assaker in Tikrit **Final segmentation** Toter bei Anschlag au Moschee in Jemen Records and a second se er Nicksechnung ins Loree Luden zu zen, dann der planze finn doch der zen, dann der planze finn doch Wachauer zwichten die Zuster den der Steine Steinen der zwicht, wei zusterer inder Steine Steinen der Micksechnung des Kohl vor der Steine dasse inder steine Steine erstern Steine erstern die Präckenschafszustehter ableite, Steine Steinen Laties keinen Steinen wirden Steine ersterne Steine Steinen wirden Steine ersterne Steinen Steinen Steinen wirden Steine ersterne Steinen Steinen wirden Steine ersterne Steinen Laties keine Steinen wirden Steine ersterne Steinen Steinen Steinen wirden Steine ersterne Steinen Steinen Steinen wirden Steine ersterne Steinen Steinen





[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR, 2013.

Zurich University of Applied Sciences







Segmentation based on textual features and neural nets

Approach

• Text block clustering (semantic distance) based on [2]

Pros

Rules can be learned implicitly • Not layout dependent

Cons

• Only text can be processed

0		derungskatalog v Stammionden zur Asyl- und Ver		IN KÜRZE
Migrationspolitik stand im rum der SVP-Delegietten- ammlung am Samstag. Der nahvorstund fordert Kanto- und Ortsparteien auf, sich n Asylunterktünfte und alhälfe-Standards zu wehren.	Nerostatiz Seruranga, Diaser ertikik cin «koylnotatorium» fir ein kilt, go- einissene Zertten fir kinitelle koyl- beweber zwie die Wilderaufmitten- eigenstäufigte Gentheumfehren, Far- tapprisikent fem Bernres suge aus hormelien und die Asobeene seine offensichtigt geschrittet, werhelt die fir keine Kinfurgunge under Staterne	Nonferenze (ff: Sorialle)[E. [Skool] seizen au pouzshal and zu prostiligi, Eir dimigeter sich Officereic-turgen auf. Er net: SVP Genzieldepolitikers, sich die Sodd Konschutz zum Köhl Zu nehmen, die 2015 aus der Skos auseid, Ausfundtenden, die nach Stei Ausfundtenden, die nach Stei Ausfundtenden, die nach Stei Daussieht Zechhorhe. Im Auseur sei- fonzielek Zechhorhe. Im Auseur sei	date die Armen werder funktorerlichtig werde. Sonst sei mar osieker au weit wie zu Zeiten von Mingtenne oder die Ein- mannetes der Framewan, als aftweiteri- else Fraupes ihner Gegetätte on wegen zu seitlerbiter Ausstatung micht standba- Maune schlung eines Begern nur zu- nehenenden Distabilität in der nitheren und weiteren Nachharshaft won false	Lubyratin sicht sich entisset (volg) - Die Uchynen Muris-Louise Bauman, die zusennen mit EDP- Maisrachtlin (Ernist Markweider im Zentrum der Koschnen "Eller sohe, altreite auf Rigs der Standackennes- erter ist auf einer einer Frieder ein, erlassen für einer einer Frieder ein, nicht sich im Grossen und Gaurea aber ertrietet. Es sich for Mehr erweich, die
Schneeberger, Kerne alden gehört zu den katholisch ge- en Kantoren, in denen die SVP rrottise Whiteschichten für ach lossen hat, sich die CVF dieses Ter- richt mehr: besetzt. Hier, wo die derzeit die Nummer 2 ist, ortet sie europpotensiel. Im Nachharkan-	confinent Version mitte mone Gegenzierd des ohte Gegenzeitung verhächtledeten Schreidern sit auch des Personentreiztigkeit mit der EU- schleterwar, und Mischenechgeter- rkält in diesetz Zusammerlang, wie der Eu- Zugang zum Schlatzut auf der Fami- bernachzug, seien zu Korragionen. Bran- en bezeichtere die agenzowahrige Aspi- ere bezeichtere die agenzowahrige Aspi-	der Attel der an Affah stamterden Somthilthesiger ohne Assochiste, im Arbeitsmeckt zwischen 2007 und 2213 von 5 un 15 Prozent ungemeichneilt. Initiative für die Armee? Prinsipesident Brutzer nahm sich Stellung zum Scheitzer der Armee-	mieten Staat bis nach Griecherland, Körne dis Schweit (zu Schweit zu Schweit zu Under Schweit zu Schweit zu Schweit zu Willersten und em pertremetissenen Weg ucht durcherben sein, wäre über eine Velderinntwich dahr mechanden- ken, unge der Vertreitigungenitzter. Abge under vansch der wannitzur	Unterlagen aus der Situting einer Parla- maniskommission im ihren Kunden aus
Ndouddan ist sie stiertrenstittiste i, und den stellt sie den einigen malent. In Obwahlen will sie Kan- en für heide Sitte in den eidgenöte en Räten, attistellen, für jenen im krut, der nuch dem Räckvitt von Hen (Alther Universit und försinnen.)	politik vor der und S01 Excipiertan de sohllig verfehrte. Der Zerenhonsstattet der Partei Sordert die Kantonal- und Ormalstienen auf, Widesstund gegen neue Asylzeumen und die Zufellung neuer Asylzeumen und Hummer Julen- chie Sommersen werd Hummer Julen-	reform im Sotionalist. Er sacholt die Schuld dahte den Mitteparteien und der FDP 2r. Sie hätten sich an den Benötin- gen vom Kutteran is dar Sieherheitspeli- tischen Kommission vor zubschliedet, der ein jährliches Minimallvalaget von 5 Mil- henden Pernken werscheiten habe. Über	direct sin Releas von «Waltoserhei- Herausgeber und SVP Nation/inte- kandidar Roger Köppel, der dafür plä- dierne, die Granner zu schlissen, um der von Wirschrzuffürfelingen domi- mierten «Vilkerwanderaus» ein Einle- ruhernetun um diek sonieten schueiten.	Moler Mithel für die Holksernes (sels) - Die Heiltenense hat übren Ertrag 2014 ern gei 10 Millionen auf krapp 183 Millionen Froderz geweigen. Gebösam- rechtekonen brachten der sozial lätigen brekerne nund die Milleuren. Prodern



w(t) w(t)

swiss group for artificial intellige

