

I primi trent'anni del terzo millennio sulle tracce di Guido

Pierre Drap

Abstract: I met Guido thirty-seven days before the start of the third millennium, after presenting at the VAST congress in Arezzo in November 2000. Our collaboration was born in the cold of a parking lot, with the attempt to start my car. A meal in Florence consolidated that bond which then evolved over time. Through rapid technological advances such as photogrammetry and artificial intelligence, we have maintained focus on their practical application in archaeology. After more than twenty years, Guido's team in Florence has achieved autonomy in photogrammetry, while in Marseille we have deepened technical discussions on archaeology. That article aims to explore technological evolution and its impact on archaeology, while also predicting future changes in the field.

Il terzo millennio è nato trentasette giorni dopo il mio incontro con Guido.

Il 24 novembre 2000 ad Arezzo, durante il congresso VAST, Guido e il suo giovane dottorando Michele sono venuti da me dopo la mia presentazione. Questo è stato l'inizio di una lunga storia. Faceva freddo e la mia auto aveva difficoltà a funzionare in queste zone poco clementi. La nostra prima collaborazione consisteva nel cercare di far partire il mio veicolo frustrato nel parcheggio gelido, oscuro e poco invitante del centro conferenze. Venivo dal sud. Marsiglia è nettamente a sud di Firenze come ogni buon cartografo sa. Quindi, pochi passi per le strade di Firenze e un buon pasto da Sabatino sono bastati per sigillare una lunga e fruttuosa collaborazione.

Fondata sulla setosa stoffa delle relazioni umane, si è sviluppata nel corso degli avanzamenti tecnologici: computer portatili, fotogrammetria, computer vision, elaborazione delle immagini, realtà virtuale e aumentata, intelligenza artificiale, ontologie e ragionatori, quindi apprendimento profondo e reti neurali.

Tuttavia, il frutto di questa collaborazione, pur nutrendosi di avanzamenti tecnologici, non ha mai perso di vista il vero scopo della sua esistenza, ovvero: uno strumento per gli archeologi. Strumenti per leggere e comprendere meglio l'edificio. E se questa collaborazione ha funzionato, e con modestia, ha funzionato, è principalmente perché si basa su una vera condivisione di conoscenze. Dopo più di vent'anni di collaborazione, infatti, il team fiorentino è oggi completamente autonomo nella fotogrammetria. Questo strumento è diventato per

Pierre Drap, CNRS, Italy, pierre.drap@gmail.com, 0000-0003-0528-9280

Referee List (DOI 10.36253/fup_referee_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Pierre Drap, *I primi trent'anni del terzo millennio sulle tracce di Guido*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0376-0.19, in Michele Nucciotti, Elisa Pruno (edited by), *Florentia. Studi di archeologia. Vol. 5 - Numero speciale - Studi in onore di Guido Vannini*, pp. 253-262, 2024, published by Firenze University Press, ISBN 979-12-215-0376-0, DOI 10.36253/979-12-215-0376-0

loro completamente ovvio. Un po' come erano la matita HB e il filo a piombo degli archeologi del millennio precedente.

Parallelamente, gli informatici di Marsiglia discutono animatamente la posizione, la forma, la rappresentazione e persino l'esistenza delle USM negative nell'ontologia che descrive il corpus archeologico in questione. Le conoscenze si mescolano nel corso degli anni di collaborazione. Il punto saliente, oltre a questi scambi produttivi di conoscenze e formalizzazioni di sapere, che affronteremo in questo articolo tratta dell'evoluzione della tecnologia e del suo impatto sui metodi di lavoro degli archeologi. Affronteremo infatti alcuni aspetti importanti dell'evoluzione tecnologica in uso all'inizio di questo millennio. Costateremo i cambiamenti avvenuti e quelli potenzialmente a venire che questi processi possono indurre nello studio e nella comprensione dell'edificio.

Discuteremo qui dell'evoluzione della fotogrammetria, dell'utilizzo delle ontologie come mezzo per esprimere la conoscenza, della realtà virtuale e aumentata per mettere in scena i risultati archeologici e dell'evoluzione del sito studiato nel tempo. Concluderemo la nostra esposizione sfiorando ciò che *propporranno i prossimi dieci anni*.

1. L'evoluzione della fotogrammetria

Dalla sua invenzione, la fotogrammetria è stata uno strumento prezioso per gli archeologi, soprattutto per gli archeologi degli elevati, in particolare per quattro delle sue molte proprietà:

- Precisione: la fotogrammetria consente di catturare dettagli con grande precisione, grazie a un rapporto di scala scelto dall'utente, in base alla focale dell'apparecchio fotografico, alla dimensione del sensore e alla distanza di ripresa.
- Non invasività: gli archeologi delle costruzioni possono creare modelli 3D dettagliati a partire da fotografie senza dover toccare o danneggiare le strutture storiche che stanno studiando. Questo processo è cruciale per la conservazione del patrimonio culturale.
- Rapidità: la fotogrammetria può essere utilizzata per creare modelli 3D in un tempo relativamente breve, il che è particolarmente utile quando gli archeologi hanno bisogno di dati precisi rapidamente. Il tempo di elaborazione del rilevamento è 'tempo macchina' non eseguito sul sito. Inoltre, i modelli 3D possono essere studiati su computer e non richiedono la presenza dell'esperto sul sito. Ciò, naturalmente, rimane discutibile e dipende dalla scala del rilevamento, dalla sua precisione, dalle capacità tecnologiche di osservazione del modello 3D ottenuto (schermo, realtà virtuale, ecc.).
- Analisi dettagliata: i modelli 3D creati dalla fotogrammetria possono essere utilizzati per effettuare analisi dettagliate degli edifici studiati, come la lettura stratigrafica, l'analisi della struttura, la rilevazione di crepe e/o danni o addirittura simulazioni di modifiche architettoniche.

Il punto di svolta dell'evoluzione della fotogrammetria si situa intorno al 2000 con l'introduzione dell'algoritmo SIFT.

SIFT (Scale-Invariant Feature Transform) è un algoritmo per la rilevazione e la descrizione di punti di interesse in immagini digitali. È stato introdotto nel 1999 da David Lowe ed è stato ampiamente utilizzato nella fotogrammetria per la creazione di modelli 3D a partire da fotografie.

Prima dell'introduzione di SIFT, la fotogrammetria si basava sulla triangolazione per determinare la posizione dei punti nello spazio utilizzando punti corrispondenti in diverse immagini. Tuttavia, questo metodo aveva dei limiti poiché i punti corrispondenti dovevano essere identificati manualmente. Questa operazione era noiosa e spesso imprecisa.

Con l'introduzione di SIFT, la fotogrammetria ha subito una grande trasformazione. Oggi, l'algoritmo permette di identificare automaticamente punti di interesse nelle immagini, come angoli e bordi, indipendentemente dalla loro scala o orientamento. Ciò consente di trovare corrispondenze tra i punti di diverse immagini, semplificando notevolmente la creazione di modelli 3D.

Utilizzando SIFT, è possibile creare modelli 3D a partire da diverse immagini utilizzando punti di interesse comuni in queste immagini. I punti di interesse vengono rilevati e descritti, quindi accoppiati per creare una corrispondenza tra le immagini. Utilizzando questa corrispondenza, è possibile determinare le posizioni 3D dei punti di interesse.

Così, SIFT ha notevolmente migliorato la precisione e l'affidabilità della fotogrammetria consentendo la rilevazione e la corrispondenza automatica di punti di interesse nelle immagini, il che ha permesso di creare modelli 3D più precisi e dettagliati a partire dalle fotografie.

Questo cambiamento di paradigma proposto da David Lowe ha aperto la strada alla fotogrammetria moderna e automatizzata. Ha consolidato il ponte tra le comunità di scienziati della fotogrammetria e di questa nuova e moderna disciplina che era la Visione Artificiale.

Si trattava quindi di abbandonare la pertinenza semantica della scelta del punto di cui si volevano trovare gli omologhi nelle altre immagini a favore dell'ottenimento di migliaia di punti omologhi non scelti dall'operatore.

2. Il numero per la semantica

La proposta di David Lowe è ora unanimemente utilizzata. Mentre aveva cercato di mantenere l'esclusività brevettando il suo algoritmo, oggi molti altri approcci simili sono utilizzati in tutti i software presenti sul mercato.

La fotogrammetria è ora uno dei tanti strumenti disponibili nella cassetta degli attrezzi degli archeologi fiorentini: avevano già una macchina fotografica, un metro, un computer. Aggiungendo solo un'altra corda al loro arco, le fotografie diventano in 3D.

Tuttavia, il cambiamento più cruciale non è certamente l'introduzione della fotogrammetria come strumento di misurazione senza contatto o come stru-

mento che consente l'analisi successiva dell'oggetto studiato, ma l'introduzione del mondo 3D nel pensiero archeologico.

Gli archeologi operano in uno spazio a quattro dimensioni (tre per lo spazio e una per il tempo), ma lo rappresentano in due dimensioni.

L'uso di solo due dimensioni per rappresentare questa complessità è dovuto principalmente a tre cause:

- Accessibilità dei dati: in molti casi, i piani 2D sono più facilmente accessibili dei modelli 3D. I piani possono essere conservati sotto forma di disegni su carta, file digitali o fotografie, il che consente un accesso rapido e facile ai dati. I modelli 3D, d'altra parte, spesso richiedono software e hardware specializzati per essere consultati e utilizzati.
- Costo: la creazione di modelli 3D è spesso più costosa della creazione di piani 2D in termini di tempo, manodopera ed attrezzature. La creazione di piani 2D può essere una soluzione più economica, ma spesso questo è un inganno. Infatti, l'ottenimento di un buon rilievo 2D passa spesso attraverso la creazione di un modello 3D in precedenza.
- Complessità: i modelli 3D possono essere molto complessi e difficili da interpretare per le persone non familiari con il software di modellazione 3D. I piani 2D, al contrario, sono più semplici e più facili da leggere e capire poiché si basano su una lunga tradizione e una conoscenza condivisa. Le conoscenze legate alla lettura dei piani sono molto più diffuse di quelle legate alla lettura dei modelli 3D, che richiedono computer, software ed esperienze in strumenti recenti e in continua evoluzione.

La famosa collaborazione di cui si parla qui è stata in grado di superare tutti questi ostacoli e molti altri: la fotogrammetria è ora uno strumento indispensabile per il rilievo, la produzione di modelli 3D è diventata una fase normale dello scavo o della lettura stratigrafica delle murature.

3. Le ontologie

Oltre alla produzione di modelli 3D, il vero problema rimane la semantica. Ottenere alcuni milioni di punti 3D, o anche una mesh densa e texturizzata, può avere alcuni vantaggi: una visualizzazione globale del sito, una visualizzazione del sito in diversi momenti significativi durante gli scavi annuali, ma il maggiore interesse di un rilievo è il legame tra le nuvole di punti 3D e la semantica degli oggetti osservati.

Ciò comporta due importanti linee di ricerca: la segmentazione delle nuvole di punti e la rappresentazione delle conoscenze necessarie per l'elaborazione di un corpus di concetti utilizzati nello studio dell'edificio.

Gli aspetti della segmentazione delle nuvole di punti, in 3D o 2D, saranno affrontati nella conclusione di questo articolo. Al contrario, un corpus di concetti relativi alle Unità Stratigrafiche Murarie è stato sviluppato e formalizzato mediante ontologie.

Le ontologie sono strutture di dati che consentono di definire relazioni tra concetti, termini ed entità di un particolare dominio. Sono utili per organizzare i dati in modo coerente e per facilitarne la comprensione e l'interpretazione.

Qui, le ontologie sono state utilizzate per consolidare le conoscenze di ciascun team: un'ontologia che descrive il processo fotogrammetrico e tutti i concetti coinvolti (punti 3D, 2D, fotocamera, calibrazione, proiezione, allineamento dei fasci) viene allineata con un'ontologia che descrive i componenti architettonici (blocco di pietra, legante, ecc.) e i concetti e le relazioni che collegano le Unità Stratigrafiche come rappresentate nella matrice di Harris.

È quindi possibile generare rappresentazioni di questi dati in 2D, come nella Figura 1, o in 3D in un processo di realtà virtuale come vedremo nella prossima sezione. Lo sviluppo di queste ontologie, la definizione precisa dei concetti in gioco come le USM negative, ad esempio, sono stati possibili solo grazie a lunghe serate di discussioni accese ma fruttuose e i cui risultati sono costantemente messi in discussione.

4. La realtà virtuale

Le nostre prime esperienze comuni in realtà virtuale risalgono al 2006. È stato con il nostro amico Paul Chapman (Glasgow School of Art · School of Simulation and Visualization) che abbiamo fatto il primo collegamento tra fotogrammetria, segmentazione dei blocchi e realtà virtuale.

La tecnologia non era ancora pronta, ma Guido, un visionario, ci spingeva in questa direzione.

Le visite virtuali dei siti, possibili in realtà virtuale e già presentate, ad esempio al congresso ICHAJ a Firenze nel 2019, avevano solo un interesse anedddotico, vagamente turistico e privo di componenti scientifiche dal punto di vista archeologico.

È solo nel 2021 che questa tecnologia si è rivelata davvero utile e utilizzata nel team. I sistemi come Oculus, senza cavi che li collegano a un computer e abbinati a un potente modellatore come Unity, consentono un facile utilizzo da parte di un utente non informatico.

Se le prestazioni grafiche non consentono ancora una reale lettura stratigrafica nel mondo virtuale ricostruito e se effettivamente il gemello digitale è ancora piuttosto opaco, è comunque sufficientemente performante per supportare interazioni sugli elementi aggiunti. È infatti possibile visualizzare contemporaneamente il modello ricostruito tramite fotogrammetria e il risultato delle letture stratigrafiche (USM e loro relazioni).

È anche possibile rappresentare diversi modelli 3D corrispondenti a diverse campagne di scavo che si susseguono negli anni. Ciò è reso possibile grazie al rigore della rilevazione 3D effettuata dal team di archeologi. Coprendo in modo esaustivo gli scavi con la fotogrammetria, viene creato un sistema di riferimento unico sull'intero sito, che consente di stimare facilmente le evoluzioni temporali del sito sotto l'effetto (certamente distruttivo) degli scavi.

5. Il terzo millennio

Negli ultimi vent'anni il mondo della tecnologia ha subito profondi cambiamenti. Ma questo è stato solo l'inizio di veri e propri rivolgimenti. Infatti, i cambiamenti avvenuti sono stati solo perfezionamenti, miglioramenti talvolta sottili e raffinati ma che conservano sempre gli stessi fondamenti.

Per ricordare, in fotogrammetria sono emersi: i descrittori SIFT, che producono migliaia di punti omologhi, lo sviluppo del *bundle adjustment*, che consente l'orientamento di migliaia di fotografie, il *dense matching* che consente il calcolo di nuvole di punti densi, ecc. Tutte queste tecniche, se hanno cambiato la forma democratizzata dell'uso della fotogrammetria, non ne hanno cambiato la sostanza intrinseca: si tratta sempre di 'semplici' calcoli di intersezioni di linee nello spazio.

Invece, gli anni a venire porteranno cambiamenti di paradigmi molto più importanti, principalmente attraverso l'apprendimento profondo, comunemente chiamato in inglese *deep learning*.

Il *deep learning* è una tecnica di apprendimento automatico che consente ai computer di imparare a riconoscere modelli in dati complessi e di prendere decisioni in base a questi modelli. Si basa su reti neurali artificiali che imitano il funzionamento del cervello umano.

Le reti neurali artificiali sono costituite da strati di neuroni interconnessi, ogni strato effettua calcoli sui dati di input e trasmette i risultati al successivo. Ogni neurone è responsabile della presa di decisione per una parte specifica del compito di classificazione o di previsione.

Il processo di apprendimento avviene in due fasi principali: l'allenamento e l'inferenza. L'allenamento consiste nel presentare alla rete neurale grandi quantità di dati etichettati, ovvero dati per i quali conosciamo i risultati attesi. La rete neurale, quindi, regola i propri parametri interni per minimizzare la differenza tra i risultati previsti e quelli reali.

Una volta che la rete neurale è stata addestrata, può essere utilizzata per la fase di inferenza. In questo momento, è in grado di prendere decisioni su dati che non ha mai visto prima. Ad esempio, una rete neurale addestrata su immagini di cani e automobili potrebbe identificare e classificare: 'Il mio cane nella mia auto'. Oltre a queste applicazioni diffuse sui social network, utilizzeremo questa tecnica sul corpus di oggetti studiato.

Le applicazioni del *deep learning* sono molto ampie, nel contesto della nostra collaborazione il *deep learning* si infiltrerà in quasi tutte le fasi chiave. La prima, quella che stiamo implementando e mettendo a disposizione, è la segmentazione delle immagini al fine di estrarre automaticamente i contorni dei blocchi di pietra. Al momento in cui scrivo queste righe, siamo nella fase di apprendimento delle reti neurali convoluzionali. Gli archeologi definiscono manualmente i contorni rilevanti dei blocchi sulle foto e poi verrà implementato *Detectron2* (una libreria di visione artificiale open source sviluppata da *Facebook AI Research* (FAIR)) per la rilevazione degli oggetti, la segmentazione semantica e altre attività di visione artificiale.

La segmentazione semantica delle immagini consentirà di automatizzare il rilievo pietra a pietra e di effettuare questo processo su migliaia di fotografie esistenti. L'automazione consentirà quindi una grande ridondanza. Infatti, finora il rilievo pietra a pietra veniva effettuato solo su una singola immagine, generalmente un'ortofoto perché il costo dell'operazione manuale tende naturalmente a voler evitare di misurare più volte lo stesso blocco. Al contrario, un processo automatico effettuato su tutte le foto conduce naturalmente a una grande ridondanza (ci sono inevitabilmente diverse foto su cui sono visibili gli stessi blocchi). Questa ridondanza di blocchi visti su diverse foto (e orientati dalla fotogrammetria) consentirà di affinare il perimetro del blocco in 3D e di garantirne l'identità.

Il secondo aspetto che sconvolgerà gli approcci del rilevamento che abbiamo messo in atto all'inizio di questo millennio è la fotogrammetria stessa o almeno il calcolo del modello 3D dalle fotocamere orientate.

Il primo passo di un processo di fotogrammetria consiste nella definizione, spesso simultanea, dei parametri intrinseci delle fotocamere (cioè ciò che le caratterizza, distanza focale, distorsione, ecc.) e dei parametri esterni, comunemente chiamati posa o orientamento, ovvero la loro posizione nello spazio.

Una volta superato questo primo passo, si tratta di calcolare il modello 3D. I punti 3D vengono calcolati a partire dai punti 2D omologhi osservati su fotografie orientate.

Questi punti 3D sono calcolati tramite triangolazione e la loro precisione dipende dall'angolo formato dalle loro *linee di vista* e dalla precisione con cui i punti su diverse foto vengono identificati come omologhi.

Negli ultimi anni è stata proposta un approccio fondamentalmente diverso, si tratta di NERF (Mildenhall et al. 2020; Martin-Brualla et al. 2020; Schwarz et al. 2020; Yu et al. 2021).

NERF (Neural Radiance Fields) è un metodo di sintesi di immagini 3D che utilizza reti neurali per approssimare la funzione di radianza della scena a partire da foto scattate da diverse angolazioni di vista. Questo metodo consente di generare immagini realistiche da prospettive che non sono state catturate direttamente dalle foto originali.

Nel 2022, il lavoro di (Muller et al. 2022) ha apportato miglioramenti significativi al metodo NERF proponendo un metodo di addestramento da un solo punto di vista, un metodo di campionamento adattativo per generare immagini più precise e un metodo di ottimizzazione dei campioni per accelerare il tempo di generazione di immagini. Questi miglioramenti hanno reso il metodo NERF più accessibile ed efficace per la generazione di immagini 3D realistiche da foto.

Nei prossimi anni, questi approcci che combinano la fotogrammetria per il calcolo dei parametri intrinseci ed estrinseci delle fotografie con metodi emergenti come NERF apriranno la strada a applicazioni di realtà virtuale che consentiranno un'immersione realistica e, infine, consentiranno di spostare il lavoro di analisi degli edifici in un mondo virtuale ricostruito in laboratorio.

Anche se già sento sorgere le obiezioni che sostengono che solo la vicinanza fisica, il tocco, il contatto con l'oggetto possono consentire un'analisi dettagliata, so che il gruppo cresciuto intorno a Guido Vannini, come ha fatto negli

ultimi vent'anni, seguirà questa strada di modernità e la seguirà con pertinenza, sapendo mettere da parte le illusioni e cogliere le opportunità tecnologiche.

Infine, il mondo dell'IA e del deep learning ci promette cose sorprendenti di cui sarebbe bene diffidare, l'apprendimento profondo e le tecniche di tipo GAN, di cui un bell'esempio è disponibile qui: <<https://this-person-does-not-exist.com>>.

Questo sito utilizza una tecnica chiamata 'Generative Adversarial Networks' (GAN) per creare immagini di volti umani realistici che non esistono realmente.

Le GAN sono un tipo di rete neurale artificiale che consente di generare immagini a partire da un insieme di dati di addestramento. I GAN sono composti da due reti neurali in competizione: un generatore e un discriminatore. Il generatore prende in input un vettore di numeri casuali (chiamato 'rumore') e produce un'immagine che viene sottoposta al discriminatore. Il discriminatore prende in input un'immagine reale proveniente dall'insieme di dati di addestramento o un'immagine generata dal generatore e cerca di determinare se è reale o generata.

Con l'apprendimento, il generatore impara a produrre immagini che ingannano il discriminatore facendole passare per immagini reali, mentre il discriminatore impara a diventare più preciso nel distinguere le immagini reali dalle immagini generate.

Le ultime innovazioni di questo tipo di approccio sono proposte da DALL-E per generare immagini a partire da una descrizione testuale della scena.

Ecco alcuni esempi, impressionanti, di immagini generate da DALL-E (Fig. 2).

Naturalmente, un archeologo esperto riconoscerà la falsificazione e un lavoro collaborativo, un gruppo ampio e multidisciplinare sarà in grado di utilizzare saggiamente queste prodezze tecnologiche e di porre le giuste domande che fanno progredire ognuno.

Grazie, Guido, per la tua incondizionata fiducia fin dall'inizio di questa storia.

Riferimenti bibliografici¹

Brown et al. 2020. "GPT-3: Language Models are Few-Shot Learners."²

Gu et al. 2021. "NeRF--: Neural Radiance Fields Without Known Camera Parameters."³

Martin-Brualla, R., Radwan N., Sajjadi M. S. M., Barron J. T., Dosovitskiy A., and D.

Duckworth. 2020. "NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections."⁴

¹ I commenti sui riferimenti bibliografici sono stati aggiunti automaticamente da ChatGPT <<https://chat.openai.com/chat>>.

² Questa pubblicazione presenta GPT-3, un altro modello di elaborazione del linguaggio naturale sviluppato da OpenAI. Sebbene non sia una pubblicazione specifica su DALL-E, descrive come i modelli di elaborazione del linguaggio naturale possano essere utilizzati per generare immagini a partire da descrizioni testuali.

³ Questa pubblicazione estende i NeRF per funzionare senza i parametri della fotocamera noti, utilizzando tecniche per stimare la posa della fotocamera dall'immagine.

⁴ Questa pubblicazione estende i NeRF per funzionare su collezioni di foto non strutturate, utilizzando tecniche per stimare la posa della fotocamera e ricostruire una scena 3D dalle immagini.

- Mildenhall, B., Srinivasan P. P., Tancik M., Barron J. T., Ramamoorthi R., and R. Ng. 2020. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis."⁵
- Muller, T., Evans A., Schied C., and A. Keller. 2022. "Instant neural graphics primitives with a multiresolution hash encoding." *ACM Trans. Graph.* 41, 4, 102: 1-102: 15⁶.
- Radford et al. 2021. "Learning Transferable Visual Models From Natural Language Supervision."⁷
- Ramesh et al. 2021. "DALL-E: Creating Images from Text."⁸
- Schwarz, K., Liao Y., Niemeyer M., and A. Geiger. 2020. "GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis."⁹
- Vaswani et al. 2017. "Attention Is All You Need."¹⁰
- Yu, A., Ye V., Tancik M., and A. Kanazawa. 2021. "PixelNeRF: Neural Radiance Fields from One or Few Images."¹¹
- Yu, A., Ye V., Tancik M., and A. Kanazawa. 2021. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4578-587.

⁵ Questa pubblicazione introduce i NeRF e descrive il loro funzionamento. I NeRF sono un approccio per rappresentare scene 3D come campi di radianza neurale, che possono essere utilizzati per sintetizzare viste di una scena da qualsiasi angolazione.

⁶ In questa pubblicazione, gli autori presentano un nuovo approccio per la generazione di immagini da primitive geometriche semplici. Le immagini vengono generate utilizzando una tecnica chiamata 'multiresolution hash encoding', che consente di archiviare e manipolare in modo efficiente le primitive geometriche nello spazio latente.

⁷ Questa pubblicazione descrive l'architettura di DALL-E in dettaglio e spiega come sia stato addestrato utilizzando dati supervisionati. Presenta anche esempi di risultati generati dal modello.

⁸ Questa pubblicazione è la prima a presentare DALL-E e descrivere come funziona. Spiega come il modello è stato addestrato e valuta i risultati ottenuti.

⁹ Questa pubblicazione presenta una variante dei NeRF chiamata GRAF (Generative Radiance Fields), che utilizza un generatore di immagini per produrre viste della scena anziché semplicemente interpolare tra viste esistenti.

¹⁰ Questa pubblicazione descrive la tecnica 'attention layer transformation' utilizzata da DALL-E per concentrarsi sulle parti importanti dell'immagine in fase di generazione. Questa tecnica è stata originariamente sviluppata per modelli di elaborazione del linguaggio naturale, ma è stata adattata per i modelli di generazione di immagini come DALL-E.

¹¹ Questa pubblicazione esplora l'uso dei NeRF per la ricostruzione di scene 3D da una singola o poche immagini. Ciò è possibile utilizzando tecniche per stimare la profondità e la posa della fotocamera dall'immagine, il che consente di ricostruire la scena in 3D.

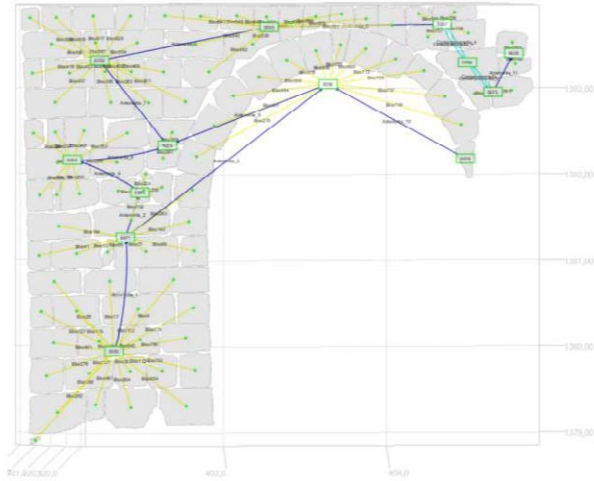


Figura 1 – Un grafico che collega le US attraverso le relazioni di Harris viene generato automaticamente sulla rappresentazione grafica degli elementi misurati.



Figura 2 – Immagini generate da DALL-E in risposta alla richiesta testuale: «Un muro fatto di blocchi squadrati, costruito dai crociati, in un castello in Giordania».